Author:
Maddrell-Mander, Samuel

Title:
An Angular Analysis of B+ K+e+e- Decays at the LHCb Experiment

An Angular Analysis of $B^+ \to K^+ e^+ e^-$ Decays at the LHCb Experiment

and Coulomb scattering correction using Generative Adversarial Networks

By

Samuel William Maddrell-Mander

Department of Physics
University of Bristol

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the School of Physics.

December 2020

Abstract

Recent anomalies in the $b \to s\ell\ell$ sector point towards possible Lepton Flavour Universality (LFU) violating effects coupling to (axial-)vector currents. An angular analysis of the differential decay rate of $B^+ \to K^+e^+e^-$ decays is used to measure the observables A_{FB} and F_H using data from pp collisions gathered at the LHCb experiment between 2011 and 2016. This dataset corresponds to a total integrated luminosity of 5fb^{-1}. The angular observables A_{FB} and F_H are only sensitive to (pseudo-)scalar and (pseudo-)tensor couplings for Standard Model (SM) like values of the observables, which therefore provides an avenue to check LFU effects and to validate the electron reconstruction at the LHCb experiment. The decay distribution is fit using a binned likelihood method, and angular observables are presented as a likelihood surface using a two dimensional Feldman Cousins construct. The result of the likelihood fit is compatible with the Standard Model prediction within 1σ. This angular analysis is the first of its kind.

Dedication and acknowledgements

First of all, thank you to my supervisor Kostas Petridis, who has demonstrated exceptional patience, encouragement, and support while guiding me to becoming a better physicist. The urgency of demanding plots was matched by an equal generosity in coffee trips. You've been a mentor, a friend, and made the last four years an absolute pleasure. (For reference to any future students, Kostas' favourite things are plots and cake, in that order.)

Thank you to Mitesh Patel and Paula Alvarez Cartelle for reading and commenting on my analysis and presentations. I am deeply indebted to Thibaud Humair for providing the R_K tuples and generous technical support, and to Dan Moise for providing more tuples over the years. Thank you to all members of the LHCb Rare Decays working group for listening when I presented my work, and providing insight and feedback on my analysis note. Special thanks to Christoph Langenbruch and Vitalii Lisovskyi for tireless reading and reviewing the analysis. Thank you to the Science and Technology Funding Council (STFC) for providing financial support throughout my studies and logistics for my LTA at CERN.

Thank you to my friends in Bristol and scattered around the country for sharing in the journey, and for waiting patiently while I finishing writing this document. Thank you to Thomas Hancock for sharing the views from the balcony in Geneva, and showing me new ways to cook. Chloé Harrison - thank you for your unwavering support, belief, and perspective. It wouldn't have been possible without you.
To my family, thank you, you gave me everything else.

Papers should be correct, talks should be interesting, and theses should be handed in.
Burton Richter

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author. Concerning the work presented in Chapter. 6 conducted in collaboration with Graphcore, the author had received an offer of employment from the company before conducting this work. Through the completion of the study and publication of the subsequent paper, the author was employed by Graphcore.

SIGNED: ..... DATE: .....

Commonly used variables

Variables related to quantities measured by the LHCb detector are mentioned throughout this note. A decay of the type $A \to b, c...$ is considered, where b and c are tracks and A is a short-lived particle candidate formed by the combination of tracks it decays into. The origin vertex of A, which is where the pp collision takes place, is referred to as the primary vertex (PV) and the decay vertex of A, defined by the interaction point of the tracks it decays into, is referred to as secondary vertex or simply decay vertex (DV).

$x_{ECAL}(b), y_{ECAL}(b)$	x and y coordinates of track b when it reaches the electromagnetic calorimeter;
$\chi^2_{DV\leftrightarrow PV}(A)$	Difference in χ^2 obtained from two fits: either assuming that all tracks come from the same point, or assuming two vertices (the primary and decay vertex). This gives an indication of the significance of the flight distance of A;
$DIRA(A)$	Directional angle: angle between the momentum vector of a particle A and the vector that links the primary vertex to the decay vertex of A;
$\chi^2_{IP}(A)$	the difference in χ^2 of a given primary vertex when reconstructed with and without the considered candidate A;
$\chi^2_{DV}(A)$	χ^2 of the fit of the decay vertex of A;
$\chi^2_{DV}(A)$/ndf	χ^2 per degree of freedom of the fit of the decay vertex of A;
$PID_\beta(b)$	Log-likelihood difference between the hypotheses that the track b is a particle of type β, and that b is a pion ($\beta = \mu, e, K$ or p). This variable uses inputs from the calorimeters and the RICH detectors;
$probNN_\beta(b)$	Probability of the track b being a particle of type β ($\beta = \mu, e, K, \pi$ or p), as estimated from information coming from all sub-detectors combined using a neural network;
$isMuon(b)$	Boolean variable that indicates whether b is compatible with being a muon, computed using information from the muon chambers;
$prob_{ghost}(b)$	Probability of b being a ghost, that is a fake track coming from a random combination of hits in the tracking system;
$\chi^2_{TrackFit}(b)$	χ^2 of the b track fit;
$nSPDHits$	Number of hits recorded in the scintillating pad detector;
$m^{track}(b, c)$	Mass of a combination of particles using only the information from the tracking systems, meaning that, in the case of electrons, the momentum from recovered bremsstrahlung is not added.
q^2	The di-lepton invariant mass squared.

Table of Contents

Chapter 1

Introduction

Over a few short decades in the 20^{th} century, a theory of fundamental particles of matter that obeyed Newtonian mechanics was completely replaced with one based on relativistic quantum field theory. This theory came to be known as the standard model (SM) of particle physics. This theory describes the constituent parts of matter, as well as the four fundamental forces that govern their interactions. If judged on the diversity and number of predictions experimentally verified, a case can be made that the SM is the closest thing to a theory of everything in modern physics. However, even a cursory glance at the deficiencies of the model show this cannot be true.

Evidence that the SM is not a fundamental theory of the underlying physics of the Universe can be found in many domains of particle physics, but is highlighted in severity on a cosmological scale. The SM cannot accommodate gravity and general relativity on the scale where such effects exist, and provides no explanation of the source of dark matter, the existence of which has been inferred from galactic observations since the 1970s[1]. More concerningly, the SM cannot explain the matter-antimatter asymmetry in the Universe. It could be argued that these comparisons are unfair, and that a true test of the SM comes on scales observed in High Energy Physics (HEP). However, even

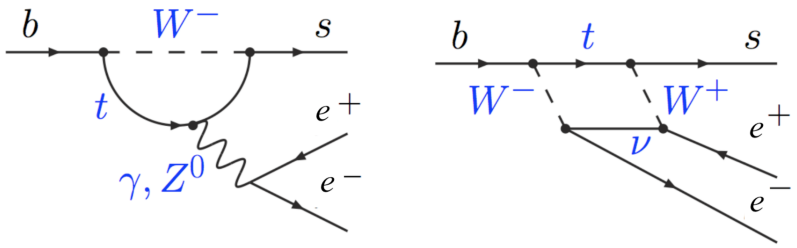

Figure 1.1: Feynman diagrams for $b \to s\ell^+\ell^-$decays in the SM at loop level.

within the confines of HEP experiments there are serious problems; these range from the naturalness and fine-tuning problems of the theory itself, to discrepancies in measured results.

The overriding conclusion is that the SM is not a fundamental theory and there must be some new physics (NP) appearing at higher energy scales that will begin to solve these problems. There are two main strategies in searching for such NP: either direct, or indirect measurements. Direct measurements look for explicit signals of NP, an approach most recently famous for the discovery of the Higgs boson in 2012 at the CMS and ATLAS experiments at CERN[2][3]. Indirect measurement experiments are used to look for evidence of NP through observations of deviations from very precise SM predictions. The NP processes are inferred in measurements of the rates and kinematic properties of decays. Such processes offer access through virtual particles to energies in great excess of direct production experiments.

Numerous recent anomalies in the $b \to sll$ flavour sector give an indication of potential lepton flavour universality (LFU) violation in (axial-)vector couplings[4, 5]. To probe these anomalies and further assumptions about LFU in other couplings, now more than ever, precise measurements of the SM properties are needed. The analysis presented in this thesis is one arm of these investigations using $B^+ \to K^+e^+e^-$ decays, from 5fb^{-1} of proton-proton collision data at the LHCb experiment [6] at CERN. This data was collected in Run1 (2011 - 2012) and in the first part of Run2 (2015 - 2016), to conduct the first angular analysis of this mode. An angular analysis of the muon mode, $B^+ \to K^+\mu^+\mu^-$, was conducted in Run1[5]; however if NP is potentially lepton flavour violating, it cannot

be assumed that these modes are identical, and the muon mode cannot be used to constrain properties of the electron mode. The decays of $B^+ \to K^+e^+e^-$are highly suppressed in the SM as there are no flavour changing neutral currents at tree level. This $b \to sll$ transition is instead dominated by electroweak penguin and box diagrams, which make it a sensitive probe of potential NP. Example Feynman diagrams shown in Fig. 1.1, describe the loop-level flavour changing neutral currents in the SM. An angular analysis provides a sensitive way to measure parameters using the kinematic distribution of the decay. In particular, the angular distribution of $b \to sll$ decays is sensitive to the Wilson coefficients for (axial-)vector, (pseudo-)scalar and tensor couplings [7].

The angular distribution can be written in terms of two observables, the forward backward asymmetry A_{FB}, and the flatness parameter F_H. Both A_{FB} and F_H are expected to be very close to zero in the SM [5]. However, this is not a general constraint. In models with NP contributions to (pseudo-)scalar or (pseudo-)tensor couplings F_H and A_{FB} may not be zero, which is particularly important for LFU violating cases where $B^0_s \to \mu^+\mu^-$ or $B^+ \to K^+\mu^+\mu^-$ cannot be used to constrain $B^+ \to K^+e^+e^-$. In this analysis, these parameters are determined using a binned maximum likelihood fit to the differential angular distribution in a $2-5$ GeV$^2/c^4$ bin of the true invariant di-electron mass squared (q^2). The choice of these bins is a result of formally accounting for resolution effects in q^2, where ${q^2}_{\text{true}}$ is the true invariant mass of the di-lepton system calculated before the effects of Final State Radiation (FSR). The angular observables are measured as averages across the ${q^2}_{\text{true}}$ bin, as in this range the observables have a flat dependence on q^2. This angular analysis provides stringent constraints on NP contributions through (pseudo-)scalar and (pseudo-)tensor couplings in the electron mode. The angular observables are only sensitive to (axial-)vector couplings for $q^2 \sim m_e^2$, and only for models beyond the SM with large (pseudo-)scalar and (pseudo-)tensor couplings. As a result, this analysis can provide vital understanding of the electron reconstruction efficiency in $b \to sll$ transitions assuming no NP in (pseudo-)scalar and (pseudo-)tensor contributions. Given the questions about lepton non-universality in $b \to sll$ decays, this angular measurement is an essential

complement to the analysis of the muon mode. The angular analysis of $B^+ \to K^+e^+e^-$ decays presented in this thesis is the first of its kind.

In addition to the angular analysis, a study was undertaken to explore potential improvements to particle track reconstruction. Generative Adversarial Networks (GANs) were used to remove the effects of Coulomb scattering from simple simulated reconstructed track hits. Extending this proof of principle to application scale could improve reconstructed track quality, which would have a direct effect on physics analyses. This reconstruction GAN was used to evaluate hardware performance as part of a study into computational approaches in next generation particle physics experiments.

This thesis is organised as follows: in Chapter 2 the Standard Model of particle physics is presented, with a focus on the flavour changing structure of the theory. The second half of the chapter focuses on $b \to s\ell^+\ell^-$decays and the unique physics motivation for studying these decays in a model-independent manner, as well as presenting key experimental results. The LHCb detector is presented in Chapter 3. In Chapter 4 the angular analysis of $B^+ \to K^+e^+e^-$decays at the LHCb detector is described in detail. This initially entails the selection of events and the removal of backgrounds before fitting a model to the invariant mass spectrum. Then, an angular model is constructed that accounts for sculpting of the true physics angular distribution due to experimental effects, with an acceptance correction. The angular model is finally fitted to the experimental data using a Feldman Cousins construct to determine the statistical confidence interval, after performing an extensive set of systematic studies to assess model uncertainty and bias. The author's contribution to upgrading the RICH mirror system is presented in Chapter 5. Finally, in Chapter 6 a preliminary investigation into the viability of using a novel GAN as a method to correct detector-level effects in particle track reconstruction is presented.

Chapter 2

Theory

In this chapter the theoretical framework of Standard Model (SM) of particle physics is introduced, some of the shortcomings of the model are discussed, and possible New Physics (NP) extensions that provide potential solutions to these problems presented. The focus of this chapter is mainly dedicated to the theoretical underpinning of electroweak decays, specifically the $b \to s\ell^+\ell^-$ transition, as this is the channel of interest in the analysis presented in Chapter 4, and the Effective Field Theory framework used to understand the analysis. There is a brief discussion of the anomalies in the $b \to s\ell^+\ell^-$ sector, and how this motivates new physics searches.

2.1 The Standard Model of Particle Physics

The SM is based on the direct product of the symmetry groups

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y. \tag{2.1}$$

Strong interactions, described by Quantum Chromodynamics (QCD), are the $SU(3)_c$ component of Eq. 2.1, while the electroweak sector is obtained through the $SU(2)_L \otimes U(1)_Y$ symmetry. The matter fields in the SM consist of three generations of spin half leptons

and quarks. The final component of the SM is the spin zero Higgs boson. The fermions consist of three generations, $i = 1, 2, 3$. Left and right handed chiral fields are given by

$$\psi_L = P_L\psi, \quad \psi_R = P_R\psi, \tag{2.2}$$

where the projection operators P_L and P_R are

$$P_L = \tfrac{1}{2}(1 - \gamma_5), \quad P_R = \tfrac{1}{2}(1 + \gamma_5), \tag{2.3}$$

respectively. Finally, γ^5 is the product of the four Dirac gamma matrices $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3\gamma^4$. The left handed quark fields can be represented as a set of three doublets

$$Q_L^i = \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix} \in \left[\begin{pmatrix} u_L \\ d_L \end{pmatrix}, \begin{pmatrix} c_L \\ s_L \end{pmatrix}, \begin{pmatrix} t_L \\ b_L \end{pmatrix} \right], \tag{2.4}$$

and the right handed quark fields as two sets of three singlets for the up and down types

$$u_R^i \in [u_R, c_R, t_R], \tag{2.5}$$

and

$$d_R^i \in [d_R, s_R, b_R]. \tag{2.6}$$

The left handed lepton fields are arranged in an identical way to the quarks

$$L_L^i = \begin{pmatrix} \nu_L^i \\ \ell_L^i \end{pmatrix} \in \left[\begin{pmatrix} \nu_L^e \\ e_L \end{pmatrix}, \begin{pmatrix} \nu_L^\mu \\ \mu_L \end{pmatrix}, \begin{pmatrix} \nu_L^\tau \\ \tau_L \end{pmatrix} \right], \tag{2.7}$$

however, as there are no right handed neutrinos in the SM the right handed leptons are described with one set of three singlets

$$\ell_R^i \in [e_R, \mu_R, \tau_R]. \tag{2.8}$$

The Lagrangian of the SM can be split into two sections as

$$\mathcal{L}_{\text{SM}} \sim \mathcal{L}_{\text{Kinetic}} + \mathcal{L}_{\text{Higgs}}; \tag{2.9}$$

where each section will be examined in turn. The kinetic component of the Lagrangian, for some field ψ, contains terms of the form

$$\mathcal{L}_{\text{Kinetic}} \sim i\bar{\psi}\gamma^\mu\mathcal{D}_\mu\psi, \tag{2.10}$$

where the gauge covariant derivative for each of the doublets is defined as:

$$\mathcal{D}_\mu^q = \left(\partial_\mu + igA_\mu^j T^j + ig_2 W_\mu^a \tau^a + ig_1 B_\mu Y\right), \tag{2.11}$$

for quarks and

$$\mathcal{D}_\mu^\ell = \left(\partial_\mu + ig_2 W_\mu^a \tau^a + ig_1 B_\mu Y\right), \tag{2.12}$$

for leptons. There are similar terms for the singlets. The second term in Eq. 2.11 contains the generators of the $SU(3)$ colour group T^j, along with the eight gluon fields A_μ^j (j = 1, 2, ... 8). The g, g_1 and g_2 are coupling constants. The third term contains the generators of the $SU(2)$ group: τ^a (defined in relation to the Pauli matrices: $\tau^a = \sigma^2/2$) and the three Electroweak gauge fields W_μ^a (a = 1, 2, 3). The fourth term has the Hypercharge generator $Y = \frac{q_Y^{L,R}}{2}$ and the single Electroweak gauge field B_μ. The second term in Eq. 2.9 gives rise to renormalisable masses in the standard model, $\mathcal{L}_{\text{Higgs}}$ is defined as:

$$\mathcal{L}_{\text{Higgs}} = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) - V(\phi^\dagger \phi) + \mathcal{L}_Y, \tag{2.13}$$

where the simplest renormalisable potential is of the form:

$$V(\phi^\dagger \phi) = \lambda(\phi^\dagger \phi)^2 - \mu^2 \phi^\dagger \phi = \lambda \left((\phi^\dagger \phi) - \frac{\mu^2}{2\lambda}\right)^2 - \frac{\mu^4}{4\lambda}. \tag{2.14}$$

The Yukawa term $\mathcal{L}_Y$ facilitates flavour structure and is explained in more detail in Eq. 2.27. The minimum of Eq. 2.14 can be found where

$$(\phi^\dagger \phi) = \frac{\mu^2}{2\lambda}. \tag{2.15}$$

The four degrees of freedom of the complex doublet can be expressed in terms of the four scalar fields ϕ_i as:

$$\frac{1}{2}(\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{\mu^2}{2\lambda}. \tag{2.16}$$

Choosing a particular direction for the minima causes the symmetry to spontaneously break. Local gauge transformations allow the vacuum expectation values (VEV) of the Goldstone bosons ϕ_1, ϕ_2, and ϕ_4 to be set to be 0. These correspond to azimuthal excitations around the potential given in Eq. 2.14, shown in Fig. 2.1, taking what is

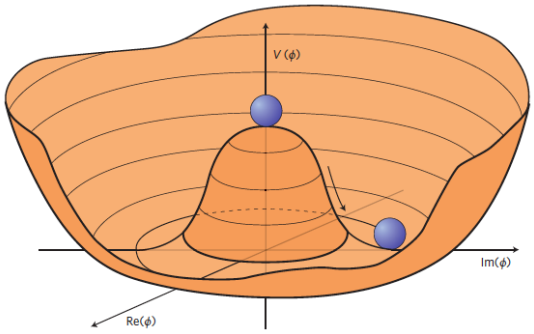

Figure 2.1: The Higgs potential is rotationally invariant.

canonically described as a wine bottle shape. The vacuum expectation value of ϕ_3 is given by:

$$\nu = \sqrt{\frac{\mu^2}{\lambda}}. \tag{2.17}$$

The vacuum expectation value has been measured experimentally as $\nu = 246.22$ GeV [3]. The symmetry is easily visualised in Fig. 2.1 where it is clear that a VEV at zero is an unstable local maxima. By choosing a specific unitary gauge where $\phi_3 = \nu + H$, ϕ can be expanded around the specific vacuum, where H is the neutral scalar Higgs field. In this unitary gauge, the symmetry is spontaneously broken and the complex doublet can be written as:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H \end{pmatrix}. \tag{2.18}$$

Through this spontaneous symmetry breaking the electroweak gauge bosons obtain mass, and the Higgs boson arises through fluctuations around the minimum of the Higgs potential.

The physical gauge fields $W^{\pm}$, Z^0 and γ in the SM do not map directly to the W^3_μ and B_μ fields in the Electroweak theory. When expanding the Higgs field kinetic term about the Higgs vacuum expectation there are two quadratic terms that lead to gauge boson masses:

$$\mathcal{L}_{\text{gauge}-\text{bosonmass}} = \frac{g_2^2\nu^2}{8}(W^{1\mu}W^1_\mu + W^{2\mu}W^2_\mu) + \frac{\nu^2}{8}(g_2W^{3\mu} - g_1B^\mu)(g_2W^3_\mu - g_1B_\mu). \tag{2.19}$$

Introducing a term for charged vector boson fields as

$$W_\mu^\pm = \frac{W_\mu^1 \pm iW_\mu^2}{\sqrt{2}} \tag{2.20}$$

and substituting into the first part of Eq. 2.19 yields a mass term for these massive charged vector bosons,

$$\frac{g_2^2\nu^2}{4}W^{+\mu}W_\mu^- = m_w^2 W^{+\mu}W_\mu^- \qquad \rightarrow \qquad m_W = \frac{g_2\nu}{2}. \tag{2.21}$$

To write the physical photon field A_μ and Z-boson field Z^0, as combinations of the neutral W_μ^3 and B_μ fields, it is convenient to define the weak mixing angle θ_w, where:

$$g_1 = \sin\theta_w\sqrt{g_1^2 + g_1^2}, \qquad\qquad g_2 = \cos\theta_w\sqrt{g_1^2 + g_1^2}. \tag{2.22}$$

Taking these definitions and substituting into the second part of Eq. 2.19, it becomes:

$$\frac{\nu^2(g_1^2 + g_2^2)}{8}(W^{3\mu}\cos\theta_w - B^\mu\sin\theta_w)(W_\mu^3\cos\theta_w - B_\mu\sin\theta_w). \tag{2.23}$$

Redefining the term in brackets as:

$$Z_\mu = W_\mu^3\cos\theta_W - B_\mu\sin\theta_W, \tag{2.24}$$

Eq. 2.23 yields the mass term for the neutral vector boson field Z^0 as

$$\frac{\nu^2(g_1^2 + g_2^2)}{8}Z^\mu Z_\mu = m_Z^2 Z^\mu Z_\mu \qquad \rightarrow \qquad m_Z = \frac{m_W}{\cos\theta_W}. \tag{2.25}$$

The final massless field that emerges is the photon as described in QED, formed as the linear combination

$$A_\mu = W_\mu^3\sin\theta_W + B_\mu\cos\theta_W. \tag{2.26}$$

In the SM all flavour violating effects are contained in the Yukawa term $\mathcal{L}_Y$ in Eq. 2.13. The Yukawa term enters the Lagrangian as

$$\mathcal{L}_{\text{Yukawa}} = \mathcal{L}_{\text{Yukawa}}^\ell + \mathcal{L}_{\text{Yukawa}}^q; \tag{2.27}$$

where the first term contains the lepton interactions and the second term the quark interactions. Consider the lepton term first, expanded as:

$$\mathcal{L}_Y^\ell = y_{ij}^\ell \overline{L}_{Li} \phi \ell_{Rj} + h.c. \tag{2.28}$$

However, in the SM there are no right handed neutrinos so this expression can be simplified to:

$$\mathcal{L}_Y^\ell = y_{ij}^\ell \overline{\ell}_{Li} \phi \ell_{Rj} + h.c., \tag{2.29}$$

where y_{ij}^ℓ is the Yukawa coupling strength between two flavour indices for the leptons. Similarly the quark Yukawa terms can be written as

$$\mathcal{L}_Y^q = y_{ij}^u \overline{Q}_{Li} \epsilon \phi u_{Rj} + y_{ij}^d \overline{Q}_{Li} \phi d_{Rj} + h.c., \tag{2.30}$$

where y_{ij}^u and y_{ij}^d are the Yukawa coupling strengths between the i^{th} and j^{th} quark generations, L and R the left and right handed chirality components, and u and d the up and down type quarks, ϵ_{ij} is the 2×2 anti-symmetric matrix[1]

$$\epsilon_{ij} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.31}$$

Substituting Eq. 2.18 into Eq. 2.29 and Eq. 2.30 the Yukawa term becomes

$$\mathcal{L}_Y^\ell = -\frac{\nu}{2} \left(\overline{\ell}_{Li} m_{ij}^\ell \ell_{Rj} + \overline{\ell}_{Ri} m_{ij}^\ell \ell_{Lj} \right) + \text{terms with } H, \tag{2.32}$$

$$\mathcal{L}_Y^q = \left(\frac{\nu}{\sqrt{2}} \right) \left(\overline{u}_{Li} m_{ij}^u u_{Rj} + \overline{d}_{Li} m_{ij}^d d_{Rj} + h.c. \right) + \text{terms with } H, \tag{2.33}$$

where $m_{ij}^{u,d}$ are the Yukawa mass matrices, defined as

$$m_{ij}^\ell = -\frac{1}{\sqrt{2}} \nu y_{ij}^\ell, \qquad m_{ij}^u = -\frac{1}{\sqrt{2}} \nu y_{ij}^u, \qquad m_{ij}^d = -\frac{1}{\sqrt{2}} \nu y_{ij}^d. \tag{2.34}$$

[1] selects the up type quarks

Moving to matrix notation, a set of six unitary 3×3 rotation matrices, $\mathbf{U}^{(u,d)^\dagger}_{\ell,L,R}$, defined by:

$$\boldsymbol{\ell}_L \to \mathbf{U}^{\ell}_{(L)} \boldsymbol{\ell}'_L \qquad \boldsymbol{\ell}_R \to \mathbf{U}^{\ell}_{(R)} \boldsymbol{\ell}'_R, \tag{2.35}$$

$$\mathbf{d}_L \to \mathbf{U}^{d}_{(L)} \mathbf{d}'_L \qquad \mathbf{d}_R \to \mathbf{U}^{d}_{(R)} \mathbf{d}'_R, \tag{2.36}$$

$$\mathbf{u}_L \to \mathbf{U}^{u}_{(L)} \mathbf{u}'_L \qquad \mathbf{u}_R \to \mathbf{U}^{u}_{(R)} \mathbf{u}'_R. \tag{2.37}$$

$$\tag{2.38}$$

can be used to diagonalise the Yukawa mass matrices as

$$\mathbf{m}^{\ell}_{\text{diag.}} = (\mathbf{U}^{\ell}_L)^\dagger \mathbf{m}^{\ell} (\mathbf{U}^{\ell}_R) = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{bmatrix}; \tag{2.39}$$

$$\mathbf{m}^{d}_{\text{diag.}} = (\mathbf{U}^{d}_L)^\dagger \mathbf{m}^{d} (\mathbf{U}^{d}_R) = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{bmatrix}; \tag{2.40}$$

$$\mathbf{m}^{u}_{\text{diag.}} = (\mathbf{U}^{u}_L)^\dagger \mathbf{m}^{u} (\mathbf{U}^{u}_R) = \frac{\sqrt{2}}{\nu} \begin{bmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{bmatrix}. \tag{2.41}$$

where $m^{\ell,u,d}_\alpha$ are the mass matrices of the leptons and quarks.

Reverting to index notation, it is convenient to express the fermion doublets in the mass basis:

$$L^i = \begin{pmatrix} \ell^i_L \\ \nu^i_L \end{pmatrix} = \begin{pmatrix} U^\ell_L \ell'_L \\ \nu_L \end{pmatrix}, \tag{2.42}$$

$$Q^i = \begin{pmatrix} u^i_L \\ d^i_L \end{pmatrix} = \begin{pmatrix} U^u_L u'_L \\ U^d_L d'_L \end{pmatrix} = U^u_L \begin{pmatrix} u'_L \\ (U^{u\dagger}_L)(U^d_L) d'_L \end{pmatrix}, \tag{2.43}$$

Now that symmetry breaking has provided masses to the constituent particles, the weak interaction term in the Lagrangian can be considered in terms of both neutral and charged

currents:

$$\mathcal{L}_{\text{interaction}} = \mathcal{L}_{NC} + \mathcal{L}_{CC}. \tag{2.44}$$

The interaction terms for quark and lepton doublets come from the $\mathcal{L}_{\text{kinetic}}$ term in Eq.2.9. The result of expanding the covariant derivatives for the doublets in Eq.2.11 and Eq.2.12, in terms of the physical gauge fields, includes terms for the charged and neutral currents for leptons and quarks. The neutral currents are mediated by the Z^μ and A^μ gauge fields, and for the left handed quark doublets the interaction terms are:

$$\mathcal{L}^Q_{NC} = i\overline{Q^i_L}\gamma^\mu i\sqrt{g_1^2+g_2^2}\left(\frac{\sigma_3}{2} - \sin^2\theta_W \cdot \mathcal{Q}\right) Z_\mu Q^i_L + i\overline{Q^i_L}\gamma^\mu ie\mathcal{Q}A_\mu Q^i_L + h.c. \quad , \tag{2.45}$$

and for the left handed lepton doublets:

$$\mathcal{L}^\ell_{NC} = i\overline{L^i_L}\gamma^\mu i\sqrt{g_1^2+g_2^2}\left(\frac{\sigma_3}{2} - \sin^2\theta_W \cdot \mathcal{Q}\right) Z_\mu L^i_L + i\overline{L^i_L}\gamma^\mu ie\mathcal{Q}A_\mu L^i_L + h.c. \quad . \tag{2.46}$$

Similarly, the charged currents are mediated by the $W^\pm$ fields and for the left handed quark doublets the interaction terms are:

$$\mathcal{L}^q_{CC} = \overline{Q}^j_L \frac{ig_2}{\sqrt{2}}\left[W^+_\mu\sigma^+ + W^-_\mu\sigma^-\right] Q^j_L, \tag{2.47}$$

and for the left handed lepton doublets:

$$\mathcal{L}^\ell_{CC} = \overline{L}^j_L \frac{ig_2}{\sqrt{2}}\left[W^+_\mu\sigma^+ + W^-_\mu\sigma^-\right] L^j_L, \tag{2.48}$$

where $W^\pm_\mu$ and Z_μ are the propagators of the charged and neural currents respectively, g_2 is the weak coupling strength, and $\mathcal{Q}$ is the electric charge generator

$$\mathcal{Q} = \frac{\sigma^3}{2} + Y, \tag{2.49}$$

and

$$\sigma^\pm = \sigma^1 \pm i\sigma^2 = \begin{bmatrix} 0 & 1\pm 1 \\ 1\mp 1 & 0 \end{bmatrix}. \tag{2.50}$$

For brevity the right handed currents are not shown, but follow a similar pattern, although there are no right handed charged-currents in the SM.

These currents can be transformed to the mass basis by substituting the rotations in Eq. 2.36 and Eq. 2.37 into Eq. 2.46 and Eq. 2.47. Shown only for quarks, the neutral current interaction results in terms including

$$\mathcal{L}_{NC} = ig_2 \overline{u}'_{iL} (U_L^u)_{i\alpha} (U_L^{u\dagger})_{\alpha j} Z_\mu \gamma^\mu u'_{jL} \qquad = \qquad ig_2 \overline{u}'_{iL} \delta_{ij} Z_\mu \gamma^\mu u'_{jL}, \tag{2.51}$$

where because the rotation matrices are unitary $(U_L^u)_{i\alpha}(U_L^{u\dagger})_{\alpha j} = \delta_{ij}$, the these currents remain unchanged. Therefore because there are no mixing terms mediated by the Z boson, there no flavour changing neutral currents in the SM at tree level, and FCNCs can only occur at loop level, making them strongly suppressed. By contrast, rotating the charged current interaction term into the mass basis yields currents at tree level as

$$\mathcal{L}_{CC} = \frac{ig_2}{\sqrt{2}} \left[W_\mu^+ \overline{u}'_{iL} (U_L^u)_{i\alpha} (U_L^{d\dagger})_{\alpha j} \gamma^\mu d'_{jL} + W_\mu^- \overline{d}'_{iL} (U_L^d)_{i\alpha} (U_L^{u\dagger})_{\alpha j} \gamma^\mu u'_{jL} \right], \tag{2.52}$$

where the mixing between up and down states can be seen in the terms $(U_L^{u\dagger})_{i\alpha}(U_L^d)_{\alpha j}$. The matrix elements $V_{ij} = (U_L^{u\dagger})_{i\alpha}(U_L^d)_{\alpha j}$ are the terms of the Cabbibo-Kobayashi-Maskawa (CKM) matrix. Expressed in the traditional form, the CKM mixing matrix is given as

$$V_{CKM} = \begin{bmatrix} c_1 & s_1 c_3 & s_1 s_3 \\ -s_1 c_2 & c_1 c_2 c_3 - s_2 s_3 e^{i\delta} & c1_{c2} s_3 + s_2 c_3 e^{i\delta} \\ -s_1 s_2 & c_1 s_2 c_3 + c_2 s_3 e^{i\delta} & c_1 s_2 s_3 - c_2 c_3 e^{i\delta} \end{bmatrix}. \tag{2.53}$$

Where $c_i = \cos\theta_i$ and $s_i = \sin\theta_i$ for flavour indices $i = 1, 2, 3$. This CKM matrix is unitary by definition, and Charge Parity (CP) symmetry is violated in the electroweak interactions by a single complex phase in the CKM matrix. Experimentally the angles θ_1, θ_2, θ_3 are found to be small, so the CKM matrix is real if $\delta = 0$, and $\delta \neq 0$ indicates CP-violation in the weak interactions. The latest global measurements of the CKM matrix can be found here [8].

2.1.1 SM Lagrangian and Free Parameter Count

The details discussed in the previous section pertain specifically to the weak interactions that are important to the $b \to s\ell\ell$ decays that are the focus of Chapter. 4. For completeness

the full Glashow-Weinberg-Salam Standard Model Lagrangian can be written as

$$\mathcal{L}_{SM} = \mathcal{L}_W + \mathcal{L}_B + \sum_{l=e,\mu,\tau} \mathcal{L}(\ell) + \sum_{l=e,\mu,\tau} + \mathcal{L}_Y(\ell) + \mathcal{L}(q) + \mathcal{L}_Y(q) + \mathcal{L}_\phi + \mathcal{L}_{QCD} + ... \quad (2.54)$$

where the ellipsis represents various gauge fixing and ghost contributions[2]. This form of the SM has been shown to be renormalisable by t'Hooft and Veltmann. As with any theory it is important to understand the number of free parameters in the model. Ignoring the quark sector there are 15 free parameters: couplings $(e(\alpha), g_1, g_2, G_e, G_\mu, G_\tau)$, masses $(M_Z, M_W, M_H, m_e, m_\mu, m_\tau)$, Higgs sector parameters (Λ, μ^2), and the weak mixing angle $\sin^2\theta_W$. The electroweak quark section introduces three angles and one complex phase from the CKM matrix, and the mass matrices m^u_{ij} and m^d_{ij} each contributing another three parameters each. This brings the running total of degrees of freedom in the $SU(3)_c \times SU(2) \times U(1)_Y$ SM to 19. From a purely theoretical point of view, that the SM contains such a degree of a freedom, while not providing a strong motivation for the particular representation of fermions and scalars, does not make it look like a compelling complete theory of everything. Such problems, combined with discrepancies in experimental results that will be discussed in the following section, motivate a search for physics beyond the SM.

[2] $\mathcal{L}_W + \mathcal{L}_B$ are the gauge invariant kinetic terms for the vector bosons

2.2 Physics Beyond the Standard Model

The standard model remains by some metrics the most successful scientific theory ever devised. However, it remains far from a fundamental theory of everything. As discussed in the previous section, the structure of the SM has problems. More glaring than just a lack of mathematical elegance, the SM simply cannot explain all experimental evidence.

- **Gravity** - In the SM there is no provision for gravity. Attempts at combining the SM with General Relativity have thus far not yielded results, and the two are suspected to be fundamentally incompatible. Additionally the weak force is some 10^{24} times stronger than gravity (at short distances).

- **Charge Parity Violation (CPV)** - The universe is matter dominated, however sources of CPV in the SM are roughly ten orders of magnitude too low to account for the matter/antimatter asymmetry observed.

- **Dark Matter / Dark Energy** - Astrophysical observations show evidence of non-interacting Dark Matter and Dark Energy that compose 95% of the energy content in the universe, neither of which are explained in the SM.

- **Neutrinos** - In the SM the neutrinos are massless, which is in direct contradiction to the observation of neutrino oscillations.

In addition to the parameter count in Sec. 2.1.1 there are some other key theoretical motivations for attempting to extend the SM.

- **Naturalness** - In addition to the large number of free parameters, there is an argument that the range of energies that the masses take is unnatural. The top quark mass is $\sim 3.6 \times 10^4$ times greater than the up quark.

- **QCD CPV** - In order to stop QCD violating C and P symmetries in the same manner as the weak force requires fine tuning parameter values to the order of 10^{-9}.

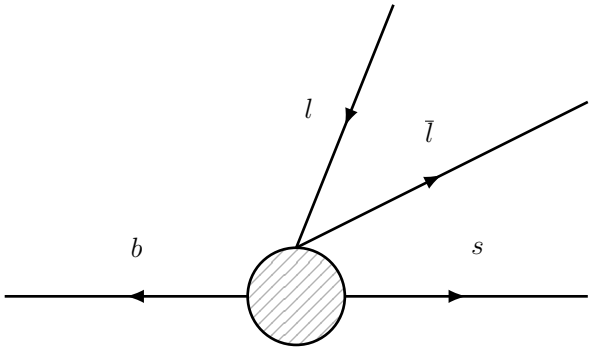

Figure 2.2: Effective Feynman diagram for the $b \to s\ell\ell$ transition.

- **Higgs mass** - Quantum corrections from higher order diagrams predict a Higgs mass on a Planck scale; that the Higgs boson is observed at the EW scale implies that to an incredi-masle degree the bare mass is cancelled by the radiative corrections. $m_H = 125.4\text{MeV} << m_{\text{Planck}} \approx 1 \times 10^{19}$.

Rare decays provide probes of physics beyond the SM as NP may not be suppressed in the same way in these channels. In particular $b \to s\ell\ell$ decays are highly suppressed in the SM as there are no flavour changing neutral currents at tree level. These $b \to s\ell\ell$ decays are instead dominated by electroweak penguin and box diagrams, which make it a sensitive probe of potential new physics.

2.3 An Effective Field Theory Approach to Model $b \to s\ell\ell$ Decays

There are two main approaches to searching for new physics; a top-down, model dependent, approach where a given theory provides an explicit prediction that can be measured; and a bottom-up approach, where generalised couplings are considered in a model independent way with no reliance on a given theory. For the later by removing the dependence on a specific type of NP model, all SM and NP couplings can be considered explicitly in their most general form. This approach also allows the separation of effects into different energy scales, low energy effects that are hard to predict can be isolated from high energy effects where there is sensitivity to NP. These bottom-up searches typically employ an Effective

Field Theory (EFT) approach where a simplified model is used to probe complex effects in a model independent way. A specific type of EFT is an Operator Product Expansion (OPE), which historically has been used to great effect, most notably by Fermi in the description of nuclear beta decay with the Fermi constant G_F. This EFT approach allowed for a description of the process as a generic four point interaction before the discovery of the $W^{\pm}$ bosons. When probing NP in $b \to s\ell\ell$ decays, a similar type of EFT is employed, as it allows the inclusion of beyond SM effects in a model independent way. In these decays the energy scale must be greater than the mass of the b-quark, which is itself much greater than the energy scale of low energy QCD effects, given as

$$\Lambda = m_b \sim 5\,\mathrm{GeV}/c^2 >> \Lambda_{QCD}. \tag{2.55}$$

Therefore, effects occurring at lower energies than Λ can be separated out. This separation comes at the cost of the introduction of hadronic form factors, as $b \to s\ell^+\ell^-$decays occur within an hadronic system. The length scale of the meson is on the order of Λ_{QCD} and therefore described by low energy QCD. These calculations are theoretically challenging, so hadronic form factors are introduced to parametrise the decay amplitudes at the meson level. These form factors are calculated using non-perturbative techniques and typically contribute the dominant source of uncertainties. Typically observables are constructed using ratios of form factors such that the uncertainties largely cancel, increasing the sensitivity to NP through the observables.

An example of how an OPE EFT can be applied to $b \to s\ell\ell$ decays is shown in Fig. 2.2 where all potential transitions are contained in the hashed circle. In a general sense, these physics models can be described as some Lagrangian where new physics enters in addition to the SM as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{SM}} + \mathcal{L}_{\mathrm{NP}}. \tag{2.56}$$

In an OPE the total Lagrangian can be expressed in an effective form as a sum of coefficients as [7]:

$$\mathcal{L}_{\mathrm{eff}} = \sum_i \mathcal{C}_i(\Lambda)\mathcal{O}_i. \tag{2.57}$$

In this effective Lagrangian, effects above the energy scale Λ (high energy, short distance) are included into the Wilson coefficients $\mathcal{C}_i$. For $b \to s\ell^+\ell^-$decays, evaluated at the b-quark mass, such short distance effects absorbed into the Wilson coefficients include the t-quark, and the Z^0 and $W^\pm$ bosons. For energies below the b-quark mass, effects are considered low energy (long distance) and are absorbed into the local operators $\mathcal{O}_i$, where each describes a type of interaction (e.g scalar, vector, axial vector, etc.). The corresponding Wilson coefficient describes the size of this contribution for both SM and NP predictions. Deviation from the SM prediction for a given Wilson coefficient indicates NP in that kind of interaction, without relying on a specific model dependant interpretation.

This effective Lagrangian contains both the SM couplings, as well as any potential NP couplings, in a model independent way as $\mathcal{C}_i = \mathcal{C}_i^{SM} + \mathcal{C}_i^{NP}$. The effective Hamiltonian for operators relevant to $b \to s\ell\ell$ decays is given by:

$$\mathcal{H}_{\text{eff}} = -\frac{4G_F}{\sqrt{2}} V_{tb}V_{ts}^* \sum_{i=7,9,10,S,P,T} \left(\mathcal{C}_i\mathcal{O}_i + \mathcal{C}_i'\mathcal{O}_i'\right), \tag{2.58}$$

where G_F is the Fermi constant, V_{tb} and V_{ts} are the CKM matrix elements. The primed operators are chiral partners that are zero or suppressed in the SM. The operators correspond to the different interaction processes, of which the relevant interaction to $b \to s\ell^+\ell^-$decays in the SM are the photon interaction described by $i = 7$; and the semi-leptonic processes described by $i = 9, 10$. These dominant operators and their chiral partners are given as as:

$$\mathcal{O}_7 = \tfrac{e}{(4\pi)^2}\overline{m}_b[\bar{s}\sigma^{\mu\nu}P_R b]F_{\mu\nu}; \quad \mathcal{O}'_7 = \tfrac{e}{(4\pi)^2}\overline{m}_b[\bar{s}\sigma^{\mu\nu}P_L b]F_{\mu\nu} \tag{2.59}$$

$$\mathcal{O}_9 = \tfrac{e^2}{(4\pi)^2}[\bar{s}\gamma_\mu P_L b][\bar{l}\gamma^\mu l]; \quad \mathcal{O}'_9 = \tfrac{e^2}{(4\pi)^2}[\bar{s}\gamma_\mu P_R b][\bar{l}\gamma^\mu l] \tag{2.60}$$

$$\mathcal{O}_{10} = \tfrac{e^2}{(4\pi)^2}[\bar{s}\gamma_\mu P_L b][\bar{l}\gamma^\mu\gamma_5 l]; \quad \mathcal{O}'_{10} = \tfrac{e^2}{(4\pi)^2}[\bar{s}\gamma_\mu P_R b][\bar{l}\gamma^\mu\gamma_5 l] \tag{2.61}$$

where $\overline{m}_b$ is the running mass of the b-quark at the given energy scale. The primed chiral partner processes are suppressed in the SM and as such are expected to be zero or close to zero. In the SM, due to the lack of right handed (axial-)vector interactions, the Wilson Coefficients for $C'^{SM}_{9,10} = 0$, and C'_7 are helicity suppressed as $C'_7 = (m_s/m_b)C_7^{SM}$.

To fully describe the scope of possible $b \to s\ell\ell$ interactions additional operators for (pseudo)-scalar and tensor couplings are required. These follow as:

$$\mathcal{O}_S^l = \tfrac{e^2}{(4\pi)^2}[\overline{s}P_L b][\bar{l}l]; \quad \mathcal{O}_S^{l\prime} = \tfrac{e^2}{(4\pi)^2}[\overline{s}P_R b][\bar{l}l]; \tag{2.62}$$

$$\mathcal{O}_P^l = \tfrac{e^2}{(4\pi)^2}[\overline{s}P_L b][\bar{l}\gamma_5 l]; \quad \mathcal{O}_P^{l\prime} = \tfrac{e^2}{(4\pi)^2}[\overline{s}P_R b][\bar{l}\gamma_5 l]; \tag{2.63}$$

$$\mathcal{O}_T^l = \tfrac{e^2}{(4\pi)^2}[\overline{s}\sigma_{\mu\nu} b][\bar{l}\sigma_{\mu\nu} l]; \quad \mathcal{O}_{T5}^l = \tfrac{e^2}{(4\pi)^2}[\overline{s}\sigma_{\mu\nu} b][\bar{l}\sigma_{\mu\nu}\gamma_5 l]. \tag{2.64}$$

Couplings are described as sums of Wilson coefficients, where the mixture depends strongly on the q^2 region that the observation is being made in. In Fig.2.3 the differential decay rate with respect to q^2 is shown, where peaks from resonances can be seen. In the SM, the non-vanishing Wilson coefficients are evaluated at the energy scale of m_W and projected to the m_b scale, given in [9] they can be summarised as

$$C_7^{SM} \sim -0.304; \quad C_9^{SM} = +4.211; \quad C_{10}^{SM} = -4.103. \tag{2.65}$$

2.4 The $B \to K\ell\ell$ Differential Decay Rate

The following derivation follows closely from that given in [7], and is used to obtain the angular distribution for $B^+ \to K^+e^+e^-$ decays given in Eq. 2.84. This angular distribution has sensitivity through two angular observables, A_{FB} and F_H, to Wilson coefficients describing (pseudo-)scalar, and (pseudo-)tensor couplings. This angular distribution is the subject of study in Chapter 4.

The full matrix element for $b \to s\ell^+\ell^-$decays can be written as

$$\begin{aligned}\mathcal{M}[\overline{B} \to K\bar{\ell}\ell] =& i\frac{G_F\alpha_e}{\sqrt{2}\pi}V_{tb}V_{ts}^*\epsilon_P(q^2)(F_V p_B^\mu[\bar{\ell}\gamma_\mu\ell] \\ &+ F_A p_B^\mu[\bar{\ell}\gamma_\mu\gamma_5\ell] + (F_S + \cos\theta_\ell F_T)[\bar{\ell}\ell] \\ &+ (F_P + \cos\theta_\ell F_{T5})[\bar{\ell}\gamma_5\ell])\end{aligned} \tag{2.66}$$

where θ_ℓ is the angle between the line of flight of the $\overline{B}$ and the ℓ^- in the di-lepton centre of mass frame, q^2 is the di-lepton invariant mass squared, and (F_V, F_A, F_{T5}, F_T, F_S, F_P)

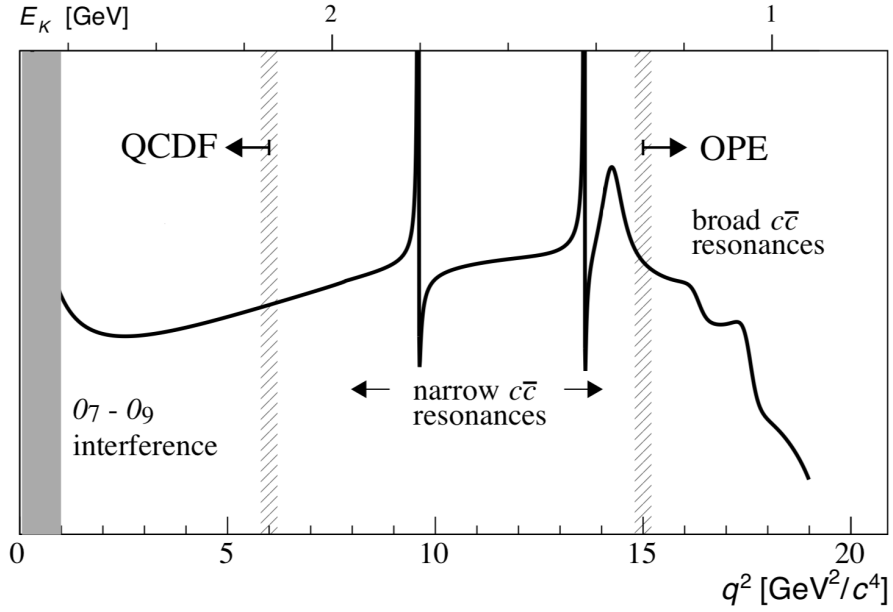

Figure 2.3: Sketch of the differential decay rate of $b \to s\ell^+\ell^-$decay as a function of q^2. The region $1.1 < q^2 < 6$ GeV^2 is dominated by the interference of $\mathcal{O}_7$ and $\mathcal{O}_9$ terms, providing excellent sensitivity to NP entering through this channel. This region is treated with QCD factorisation. The central region is dominated by the narrow $c\bar{c}$ resonances of the J/Ψ and $\psi(2s)$, where the form factor calculations are theoretically challenging. Above $q^2 > 15$ GeV^2 Operator Product Expansion methods are valid. Figure adapted from [10]. The rare mode is defined to the region $1.1 < q^2 < 6$ GeV, and the control mode in the region $6 < q^2 < 12$ GeV.

are factors containing the Wilson coefficients. The matrix element in Eq. 2.66 can be parametrised by three form factors, $f_{0,+,T}(q^2)$. The form factors contain the main source of uncertainties in this expression. These functions translate to Wilson coefficients as

$$F_A = C_{10}, \tag{2.67}$$

$$F_T = \frac{2\sqrt{\lambda}\beta_l}{M_B + M_K}\frac{f_T(q^2)}{f_+(q^2)}C_T^l, \tag{2.68}$$

$$F_{T5} = \frac{2\sqrt{\lambda}\beta_l}{M_B + M_K}\frac{f_T(q^2)}{f_+(q^2)}C_{T5}^l, \tag{2.69}$$

$$F_P = \frac{1}{2}\frac{M_B^2 - M_K^2}{m_b - m_s}\frac{f_0(q^2)}{f_+(q^2)}(C_P^l + C_P^{l'}) + m_l C_{10}\left[\frac{M_B^2 - M_K^2}{q^2}\left(\frac{f_0(q^2)}{f_+(q^2)} - 1\right) - 1\right] \tag{2.70}$$

$$F_S = \frac{1}{2}\frac{M_B^2 - M_K^2}{m_b - m_s}\frac{f_0(q^2)}{f_+(q^2)}(C_S^l + C_S^{l'}) \tag{2.71}$$

$$F_V = C_9 + \frac{2m_b T_P(q^2)}{M_B \epsilon_P(q^2)} + \frac{8m_l}{M_B + M_K}\frac{f_T(q^2)}{f_+(q^2)}C_T^l. \tag{2.72}$$

Where

$$\lambda = M_B^4 + M_K^4 + q^4 - 2(M_B^2 M_K^2 + M_B^2 q^2 + M_K^2 q^2), \qquad \beta_\ell = \sqrt{1 - 4\frac{m_l^2}{q^2}}. \tag{2.73}$$

In the SM $F_S^{SM} = F_T^{SM} = F_{T5}^{SM} = 0$. From the matrix element in Eq. 2.66, the double differential decay rate in q^2 and $\cos\theta_\ell$ for a given lepton flavour l can be expressed as a quadratic in terms of $\cos\theta_l$ as

$$\frac{d^2\Gamma}{dq^2 d\cos\theta_\ell} = a_l(q^2) + b_l(q^2)\cos\theta_\ell + c_l(q^2)\cos^2\theta_\ell. \tag{2.74}$$

Where the lepton charge asymmetry angle θ_ℓ is the opening angle between the positively (negatively) charged lepton in the di-lepton rest frame and the di-lepton system in the B^+ (B^-) rest frame. The three coefficients of the quadratic in Eq. 2.74 are given as:

$$\begin{aligned}\frac{a_l(q^2)}{\Gamma_0\sqrt{\lambda}\beta_l\epsilon_P^2} =& q^2(\beta_l^2|F_S|^2 + |F_P|^2) + \frac{\lambda}{4}(|F_A|^2 + |F_V|^2) \\ &+ 2m_l(M_B^2 - M_K^2 + q^2)Re(F_P F_A^*) + 4m_l^2 M_B^2 |F_A|^2;\end{aligned} \tag{2.75}$$

$$\begin{aligned}\frac{b_l(q^2)}{\Gamma_0\sqrt{\lambda}\beta_l\epsilon_P^2} =& 2\left\{q^2\left[\beta_l^2 Re(F_S F_T^*) + Re(F_P F_{T5}^*)\right]\right. \\ &\left. + m_l\left[\sqrt{\lambda}\beta_l Re(F_S F_V^*) + (M_B^2 - M_K^2 + q^2)Re(F_{T5}F_A^*)\right]\right\};\end{aligned} \tag{2.76}$$

$$\begin{aligned}\frac{c_l(q^2)}{\Gamma_0\sqrt{\lambda}\beta_l\epsilon_P^2} =& q^2\left(\beta_l^2|F_T|^2 + |F_{T5}|^2\right) - \frac{\lambda}{4}\beta_l^2(|F_A|^2 + |F_V|^2) \\ &+ 2m_l\sqrt{\lambda}\beta_l Re(F_T F_V^*);\end{aligned} \tag{2.77}$$

where

$$\Gamma_0 = \frac{G_F^2\alpha_e^2|V_{tb}V_{ts}^*|}{512\pi^2 M_B^3}. \tag{2.78}$$

The q^2 dependent functions in Eq.[2.75 - 2.77] have lengthy generalised forms, however, in the q^2 region where $q^2 >> m_l^2$ the limit $m_l \to 0$ can be assumed, and the functions reduce to:

$$\begin{aligned}\frac{a_l(q^2)}{\Gamma_0\sqrt{\lambda}\beta_l\epsilon_P^2} &= q^2(\beta_l^2|F_S|^2 + |F_P|^2) + \frac{\lambda}{4}(|F_A|^2 + |F_V|^2), \\ \frac{b_l(q^2)}{\Gamma_0\sqrt{\lambda}\beta_l\epsilon_P^2} &= 2\left(q^2[\beta_l^2 Re(F_S F_T*) + Re(F_P F_{T5}*)]\right) = 0, \\ \frac{c_l(q^2)}{\Gamma_0\sqrt{\lambda}\beta_l\epsilon_P^2} &= q^2(\beta_l^2|F_T|^2 + |F_{T5}|^2) - \frac{\lambda}{4}\beta_l^2(|F_A|^2 + |F_V|^2).\end{aligned} \tag{2.79}$$

In the SM it is easy to see $a_l(q^2) = -c_l(q^2)$ due to the vanishing of various F_X factors. In the SM $C_S = C_P = C_T = C_{T5} = 0$, as there are no (pseudo-)scalar and (pseudo-)tensor currents in $b \to s\ell\ell$ decays.[3] It should be noted that the boundary conditions of the phase space restrict the dilepton masses as $4m_l^2 < q^2 < (M_B - M_K)^2$. Therefore, in general analyses are keep within a $1.1 < q^2 < 6$ GeV^2 range and above $q^2 > 15$ GeV^2. This is because close to the resonances shown in Fig. 2.3 the form factor calculations break down.

The angular distribution in Eq. 2.74 can be given in terms of q^2-integrated coefficients as

$$\frac{d^2\Gamma}{dq^2 d\cos\theta_\ell} = A_l + B_l \cos\theta_\ell + C_l \cos^2\theta_\ell \tag{2.80}$$

where the q^2 integrated terms are

$$A_l = \int_{q^2{}_{min}}^{q^2{}_{max}} a_l(q^2)dq^2, \quad B_l = \int_{q^2{}_{min}}^{q^2{}_{max}} b_l(q^2)dq^2, \quad C_l = \int_{q^2{}_{min}}^{q^2{}_{max}} c_l(q^2)dq^2. \tag{2.81}$$

The decay rate and forwards-backward asymmetry can be expressed in terms of the integrated variables as

$$\Gamma_l = 2\left(A_l + \frac{1}{3}C_l\right), \qquad A_{FB}^l = \frac{B_l}{\Gamma_l}. \tag{2.82}$$

An additional variable

$$F_H = \frac{2}{\Gamma_l}(A_l + C_l), \tag{2.83}$$

known as the flatness parameter is also introduced. Both A_{FB} and F_H have the useful property that they are normalised by the decay rate, and therefore, benefit from reduced uncertainties due to cancellations between the numerator and denominator compared to the uncertainties in Γ_ℓ. Writing the angular Eq. 2.80 in terms of Eq. 2.82 and Eq. 2.83 is equivalent to

$$\frac{1}{\Gamma}\frac{d\Gamma_\ell}{d\cos\theta} = \frac{3}{4}(1 - F_H^\ell)(1 - \cos^2\theta_\ell) + \frac{1}{2}F_H^\ell + A_{FB}^\ell \cos\theta_\ell. \tag{2.84}$$

This differential decay rate is the distribution used to obtain the angular observables in the analysis presented in Chapter. 4. In the SM F_H^l and A_{FB}^ℓ vanish in the $m_\ell \to 0$ limit, resulting in a $\cos^2\theta$ angular dependence of $\overline{B} \to \overline{K}\ell\ell$ decays. Such $\overline{B} \to \overline{K}\ell\ell$ decays have

[3]The Higgs coupling to leptons is unusally small.

potential sensitivity to a number of new physics currents through the angular observables A_{FB} and F_H. The observables A_{FB} and F_H are only sensitive to $F_{A,V}$ (and consequentially $\mathcal{C}_9$ and $\mathcal{C}_{10}$) for $q^2 \sim m_\ell^2$. However, in the limit of $m_\ell \to 0$, the angular observables are sensitive to $F_{S,T,P,T5}$. The angular decay rate could have contributions from higher power $\cos^n \theta_\ell$ contributions from the weak Hamiltonian or from QED corrections, however these are suppressed. For the differential decay rate in eq. 2.84 to be physically meaningful it must be positive definite for all values of $\cos\theta_l$, however, this gives rise to un-physical regions in the A_{FB}, F_H space. These boundary conditions can be specified as:

$$\begin{aligned} 0 \leq F_H \leq 3, \\ |A_{FB}| \leq F_H/2. \end{aligned} \tag{2.85}$$

These limits tightly constrain the SM predictions which for $B^+ \to K^+e^+e^-$ decays is $F_H \sim 0$ and $A_{FB} \sim 0$. However, this is not a general constraint. In models with new physics contributions to (pseudo-)scalar or (pseudo-)tensor couplings F_H and A_{FB} may not be zero, which is particularly important for lepton universality violating (LUV) cases where $B_s^0 \to \mu^+\mu^-$ or $B^+ \to K^+\mu^+\mu^-$ cannot be used to constrain $B^+ \to K^+e^+e^-$.

2.5 R_K

Another variable that provides insight into potential new physics entering through $B \to Kll$ decays is the R_K variable. This is a measure of the ratio of $B^+ \to K^+\mu^+\mu^-$ to $B^+ \to K^+e^+e^-$ decay rates under the same q^2 selection.

$$R_K = \frac{\Gamma_\mu}{\Gamma_e} = \int_{{q^2}_{min}}^{{q^2}_{max}} \frac{d\Gamma_\mu}{dq^2} dq^2 \Bigg/ \int_{{q^2}_{min}}^{{q^2}_{max}} \frac{d\Gamma_e}{dq^2} dq^2 = \frac{\Gamma_\mu F_H^\mu - 4/3 C_\mu}{\Gamma_e}. \tag{2.86}$$

This provides a probe of lepton flavour dependent NP contributions. Additionally it can be shown that R_K and F_H are related in a model independent way[7], as

$$R_K \cdot (1 - F_H^\mu - \Delta) = 1, \tag{2.87}$$

where:

$$\Delta = \frac{3}{4}\frac{C_e - C_\mu}{\Gamma_\mu} - \frac{F_H^e}{R_K}. \tag{2.88}$$

In the SM Eq.2.88 simplifies in the limit $m_e = 0$, yielding $F_H^e = 0$ and $\Gamma_e = -4/3C_e$. Therefore in the SM, $\Delta_{SM} \sim m_\mu^2$.

As detailed in previous sections, in the SM there is a general observed property of lepton flavour universality. Many NP models suggest mechanisms to break this universality as a by product of introducing additional degrees of freedom. This ratio provides scope to probe the LFU properties of the SM, and the impact NP models may have on the assumption.

2.6 Experimental Results

Three sets of experimental results of $b \to s\ell^+\ell^-$decays will be discussed in this section. Firstly the results of angular analyses of $b \to s\ell^+\ell^-$decays from the LHCb and Belle experiments. Secondly, the results of a number of individual branching fraction measurements of $b \to s\mu\mu$ modes, and thirdly the results of two key lepton flavour universality tests. Finally an analysis of the global situation will be presented.

2.6.1 Angular observables

In Run1, the LHCb collaboration performed an angular analysis on the $B^0 \to K^{0*}\mu^+\mu^-$ decays, spanning 8 bins in q^2 from $0.1\,\text{GeV}^2$ to $19.0\,\text{GeV}^2$. A set of form factor independent angular observables are used of the form P_i', following the derivation in [11]. A deviation was found in the angular distribution across a number of parameters, and was particularly apparent in the angular observable P_5', the global fits pointed towards a deviation in the C_9 Wilson coefficient (axial-vector coupling)[12], where further details can be found. In Fig. 2.4 the SM predictions using a theoretical framework is shown against experimental results for this P_5' variable. The exact magnitude of the tension depends on the choice of theory model to describe the nuisance parameters. However, in [13], a tension to the SM of 2.5σ and 2.9σ in the 4-6 GeV and 6-8 GeV bins respectively is found. The Belle collaboration performed a corroborating measurement in 4 bins of q^2 in the same

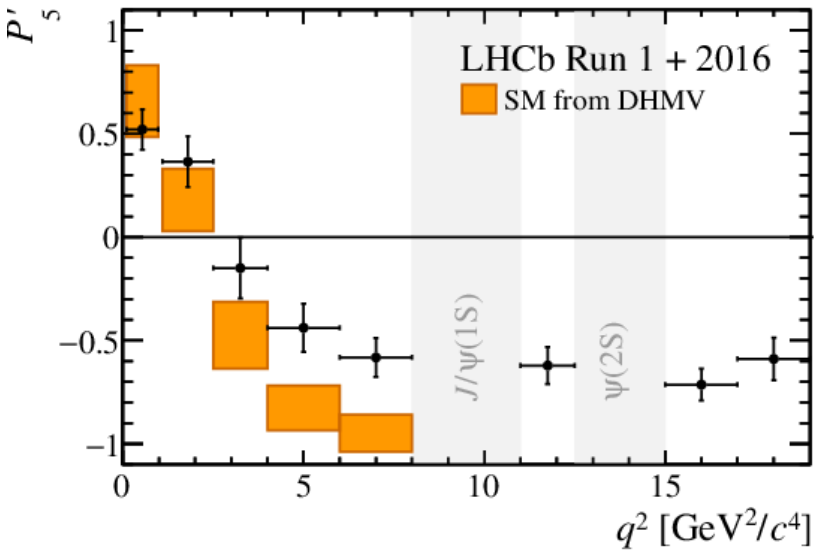

Figure 2.4: The P'_5 experimental result, studied at LHCb, Belle, etc.[12]

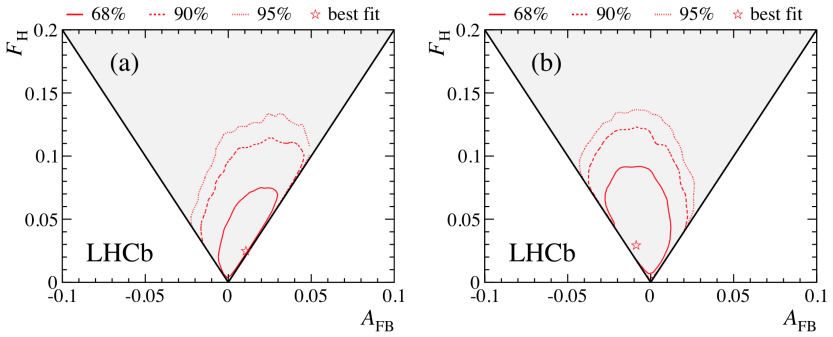

Figure 2.5: Angular analysis results of $B^+ \to K^+\mu^+\mu^-$ decays measured at the LHCb experiment, using data from Run1. Two regions of q^2 are shown, $1-6$ GeV (left) and $15-20$ GeV (right).[5]

kinematic region and found a similar tension with the SM[14].

A similar angular analysis was performed using $B^+ \to K^+\mu^+\mu^-$ decays, with sensitivity to (pseudo-)scalar and tensor amplitudes, with Run1 data at the LHCb experiment. This is the muon mode partner to the analysis presented in Chapter. 4. The amplitudes were found to be compatible with the SM predictions, as can be seen in Fig. 2.5. An updated measurement using Run2 data is currently underway. The boundaries of the physical region are shown as the grey region following the definitions in Eq. 2.85. The proximity of the best fit point with respect to the boundaries adds experimental complexity to the measurement.

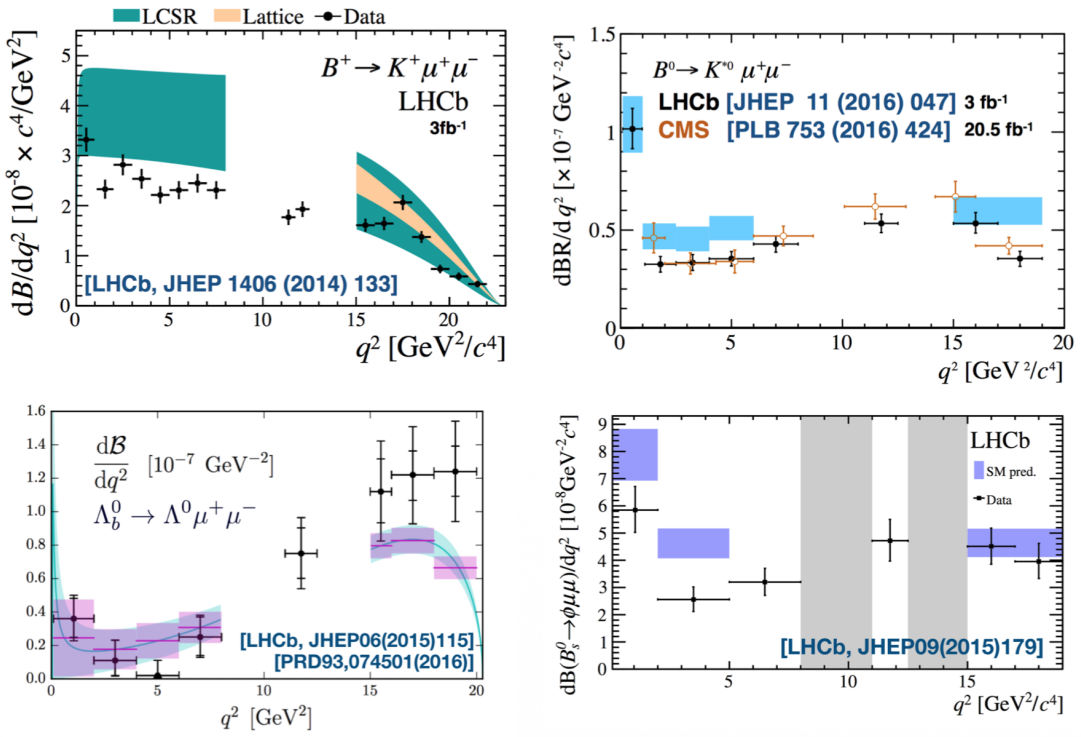

Figure 2.6: (Top Left) Branching fraction results from the LHCb experiment for $B^+ \to K^+\mu^+\mu^-$decays shown against theory predictions. (Top Right) Branching fraction results for $B^0 \to K^0\mu^+\mu^-$ decays from the LHCb and CMS experiments, shown against theory predictions. (Bottom Left) Branching fraction results for $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ from the LHCb experiment shown against theory predictions. (Bottom Right) Branching fractions of $B_s^0 \to \phi\mu^+\mu^-$ decays shown against theory predictions. [15–20]

2.6.2 Branching Fraction Results

Branching fraction measurements are good probes of NP as additional couplings could alter the observed rate in sensitive channels. The branching fractions in both $B^+ \to K^+\mu^+\mu^-$ and $B^0 \to K^{*0}\mu^+\mu^-$ channels have been observed as lower than predicted by theory for lower q^2 regions. Both results can be seen in Fig. 2.6. That the $B^0 \to K^{*0}\mu^+\mu^-$ tension has been independently observed at CMS paints a consistent picture of the physics. Additional discrepancies have been seen in $B_s^0 \to \phi\mu^+\mu^-$ and $\Lambda_b^0 \to \Lambda^0\mu^+\mu^-$ channels, also showing a lower than expected rate observed for muon modes. However, branching fraction measurements are susceptible to relatively large uncertainties from theory. These results are also shown in Fig. 2.6.

2.6.3 Ratio tests

Ratio tests of the form $R_{K^{(*)}}$, as detailed in the previous section 2.5, provide a sensitive probe of NP as the hadronic uncertainties are largely reduced, the QED contributions are on the order of 1%, and the experimental systematic uncertainties largely cancel. This clear advantage over branching fraction measurements make them the gold standard in searches for NP. The two most prominent results are those of R_K and R_{K^*}. The BaBar and Belle collaborations both made measurements of R_K and R_{K^*} and found them to be in agreement with the SM, however with large statistical uncertainties. In Run1 at the LHCb experiment the measurement of R_K and R_{K^*} were performed, and found to be in tension with the SM prediction. The LHCb measurement of R_K is performed as a double ratio where the branching fraction ratio is normalised by a control mode using the J/Ψ, this serves to cancel a number experimental uncertainties. The latest published result from data collected at the LHCb experiment from Run1 and the first part of Run2 is $R_K = 0.846^{+0.06}_{-0.054}$, which is at 2.5σ from the SM.[4] This clearly remains a contentious result, and as such a much needed follow up result is being performed. Measured values for R_K from numerous experiments are shown in Fig. 2.7 (left).

The analogous measurement R_{K^*} was also made at LHCb using Run1 data. A key difference to the R_K measurement is that it is performed in two q^2 regions, a low bin ($0.045 < q^2 < 1.1\,\text{GeV}^2$ that has particular sensitivity to C_7' and C_7, as well as a central q^2 bin in $1.1 - 6\,\text{GeV}^2$ which has strong sensitivity to C_9 and C_{10}. The measured values in these two regions are in tension with the SM at the order of 2.2 ad 2.5 standard deviations. A follow up Run2 measurement of these parameters is also underway.

2.7 Phenomenological Interpretation

The combination of measurements probing NP in the flavour sector is essential to understand the global picture. In Fig. 2.8 the latest global fit to the Wilson coefficients is shown that combines results from branching fraction measurements, angular analyses,

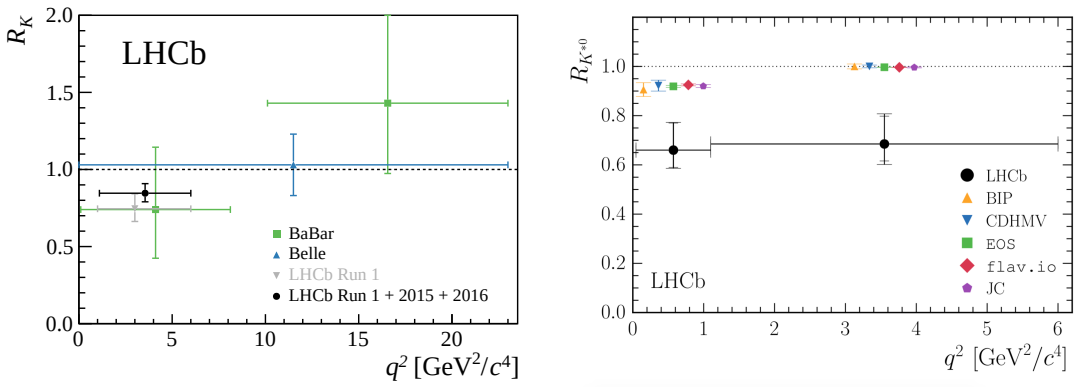

Figure 2.7: Observations of R_K from LHCb, BarBar, Belle (left) and R^*_K measurement from LHCb Run1 compared to theory predictions (right).

and ratio tests. Shown for the C_9^{NP} vector and C_{10}^{NP} axial vector couplings in $b \to s\mu\mu$ decays, a global shift is seen, indicating a preference for non-zero C_9^{NP}.The global fits show that all measurements favour a NP contribution in vector-dilepton couplings and provide a strong hint of NP in the $b \to s\ell\ell$ sector. The global tension is $> 5\sigma$.

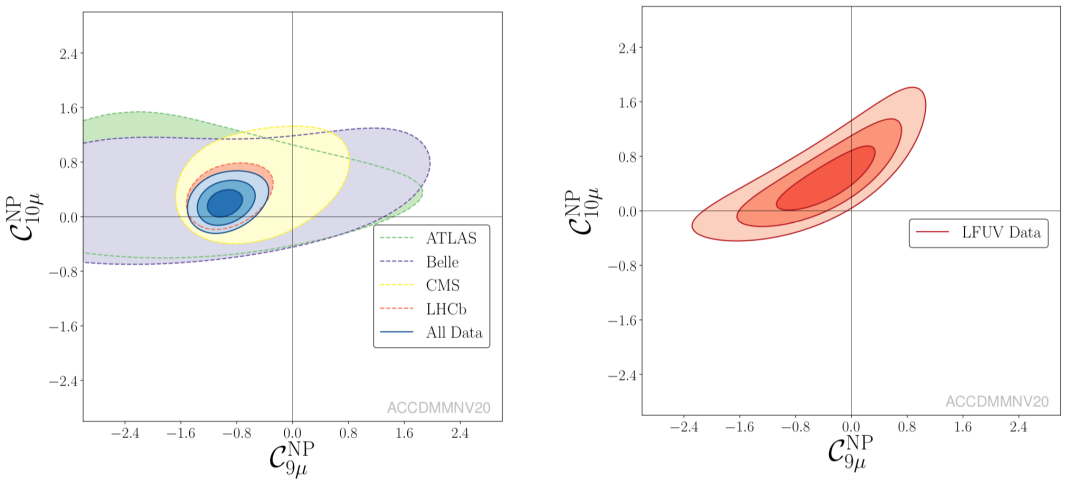

Figure 2.8: Global fits to Wilson Coefficients C_9^{NP} and C_{10}^{NP} for the $b \to s\mu\mu$ decays, combining angular measurements, branching fractions, and LFU ratio tests. [21]

2.8 Summary and future prospects

The Standard Model has been in incredibly successful theoretical framework to describe interactions of fundamental particles on high energy scales, however, as has been outlined briefly in this chapter there are a number of deficiencies that suggest a deeper theoretical

description of nature is required. This chapter also presented the pertinent phenomenological interpretation required to understand the motivation behind the analysis presented in Chapter 4.

Chapter 3

LHCb experiment at the LHC

The LHCb detector is one of four main experiments situated at the Large Hadron Collider (LHC) at European Organisation for Nuclear Research (CERN) in Geneva. The following chapter provides detail about the LHC itself, and then specifics of the LHCb detector, with focus on the most relevant areas for the work presented in this thesis.

3.1 The LHC

The LHC at CERN is a proton-proton collider housed in a 27 km long tunnel under the Swiss-French countryside outside Geneva that used to hold the Large Electron Positron (LEP) collider, an areal photograph of the region is shown in Fig. 3.1. The LHC is designed to accelerate beams of protons up to an energy of 7 TeV travelling in opposite directions. The two beams are crossed at four specific interactions points around the ring resulting in a centre of mass (COM) energy of 14 TeV. In order to reach the final beam energy of 7 TeV the protons must be stepped in energy through a series of smaller accelerators, primarily at the CERN Meyrin site. Protons are extracted from hydrogen and accelerated in the Linear Accelerator (LINAC 2) up to 50 MeV, where they enter the

first booster ring BOOSTER which takes them to 1.4 GeV, then the Proton Synchrotron (PS) to reach 26 GeV, then the Super Proton Synchrotron to reach 450 GeV before finally entering the Large Hadron Collider (LHC) to reach the final 7 TeV beam energy.

Once in the LHC, the beam diameter is restricted using a series of quadro-pole super-cooled Niobium-Titanium magnets, each with a maximum field strength of 8.3 T, which are required to bend the beam around the ring. The proton beams are not continuous; but rather consist of bunches of protons, where each bunch contains approximately $1.2 - 1.4 \times 10^{11}$ protons with a separation of 25 ns (referred to as the bunch separation), corresponding to a bunch rate of 40 MHz. This corresponds to a total of approximately 5.6×10^{18} protons delivered every second. The four interaction points (IP) around the LHC ring contain the main experiments operated at CERN. The two general purpose experiments, CMS and ATLAS, have physics programs largely focused on studying high transverse momentum (p_T) particles and search for general new physics (NP) signals using their 4π solid angle coverage. These experiments are at the forefront of making precise measurements of the Higgs boson candidate, which was discovered simultaneously at CMS and ATLAS in 2012, however also have a wide ranging b and t quark program. The b physics program in the coming years will provide a valuable and highly complimentary avenue to validate heavy flavour measurements made at LHCb with a large statistics sample, with fundamentally different systematic uncertainties.

The other main experiments are: LHCb, which will be described in more detail in the following section 3.2; and the ALICE experiment which is another general purpose experiment with a physics program focused on heavy ion physics. There are many other experiments located at CERN that cannot be listed here, but all contribute to a rich and diverse physics program at CERN.

3.2 The LHCb Physics Program

The LHCb experiment was designed to study the heavy flavour physics of beauty and charm hadrons, and to look for indirect evidence of NP through precision measurements

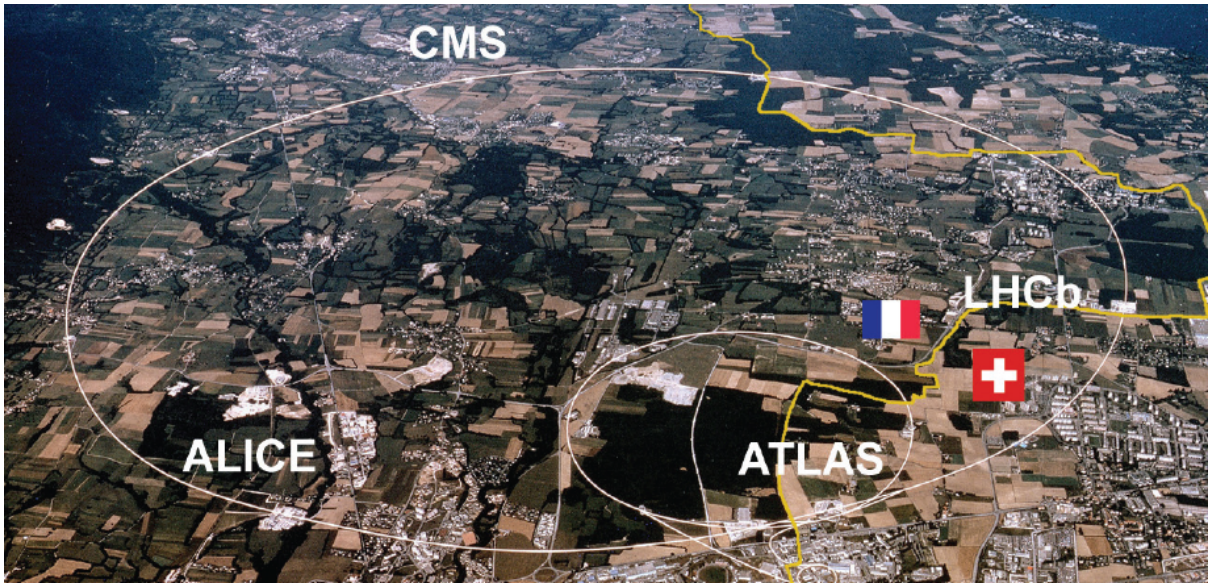

Figure 3.1: An aerial view of the Swiss-French border with the projection of the LHC main ring and accelerator rings shown. LHCb is situated just over the border from Meyrin, next to the airport and the author's former residence.

of CP violating (CPV) processes and rare decays.

The experiment was run (in Run2) at an energy of 13 TeV, corresponding to a $b\bar{b}$ cross section of 500 mb, but at a relatively modest luminosity of 2×10^{32} compared to other LHC experiments. The reason for the modest luminosity when compared to CMS or ATLAS is that each event seen by LHCb is dominated by one single pp interaction, drastically reducing occupancy in the sub detectors. This is essential as the Primary Vertex (PV) that the B-meson comes from must be identified, if there are too many PVs present in an event this becomes difficult. A secondary but very important advantage is that it reduces the radiation damage to the detector, increasing the lifetime and reducing the overall operation cost by preserving more components from run to run. These running conditions mean that 10^{12} $b\bar{b}$ pairs can be produced in what is referred to as a 'canonical' year of data taking, i.e. the data collected if the experiment was run for 1/3 of an astronomical year. In Fig.3.2 the cumulative data taking at the LHCb experiment is shown as well as the integrated luminosity compared per year. While data has been taken up to the end of 2018 when the second long shut down started, the analysis presented in Chapter 4 only uses samples from Run1 and 2015+2016.

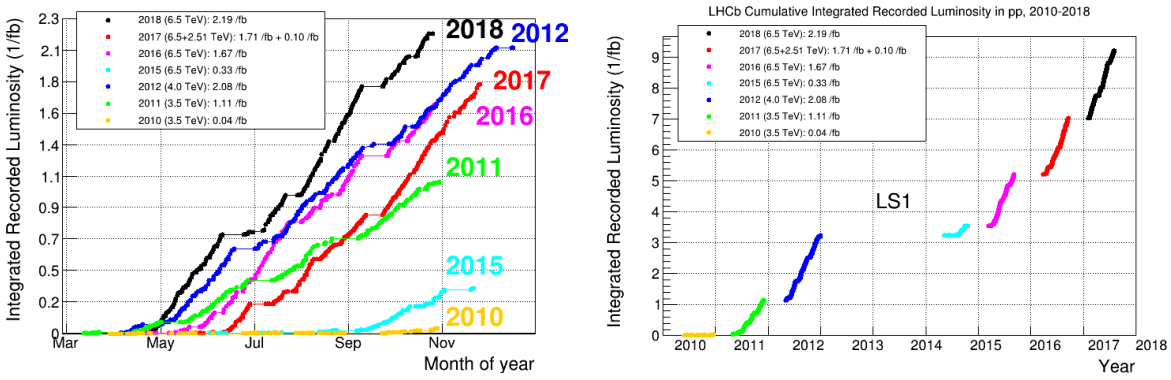

Figure 3.2: The left plot shows taking at the LHCb experiment in terms of integrated luminosity each year during 2012 - 2018. The plot on the right shows the cumulative data gathered over time.

The family of experiments known as 'b-factories' include two key competitors to the LHCb experiment, BaBar and Belle (II). Both experiments are electron-positron colliders operating at lower COM energies, and as such they have much cleaner environments with significantly reduced backgrounds. However, they have significantly smaller total datasets of b candidates. Belle II, for instance, is expected to become competitive with the current LHCb datasets only in the next five years. These two experiments are vital in the understanding of indirect flavour physics measurements, as each experiment is sensitive to subtly different uncertainties, and provide important cross checks. While the luminosity at the LHCb experiment is lower than at electron-positron experiments, the $b\bar{b}$ cross section is significantly higher. However, this results in a polluted background in the LHCb data due to the complex environment of a pp collision that must be removed for analysis.

The structure of LHCb is a forward arm spectrometer (single sided) that results in a pseudo-rapidity coverage of only $1.8 < \eta < 4.9$, where

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right), \tag{3.1}$$

and θ is the angle between the beam axis and the particle momentum. This corresponds to 4% of the total solid angle compared to the GPDs outlined in the previous section. The $b\bar{b}$ pairs are highly correlated when produced and are boosted along the beam-line due to their relatively low mass and some 25% of the candidates are captured in the geometrical acceptance. At the LHCb experiment backgrounds are greatly reduced by exploiting

the boost of the B-meson pair. The displacement observed as a result of this boost makes it possible to correctly identify daughter particles from each B-meson, reducing the combinatorial background. The forward nature of the detector also provides a practical benefit as well, the commissioning and maintenance of the detector is significantly easier than a hermetically sealed detector as much of the dead material and electronics can be routed outside the pseudo-rapidity region. This construction reduces the material budget of the detector and improving resolution of the momentum. Additionally it makes servicing the detector notably simpler than the GPDs.

The LHCb detector itself is some 20 m in length and up to 5 m tall at the furthest point of the detector, a diagram showing the scale and order of the sub-detectors is shown in Fig. 3.3. The detector is split into various sub-detectors that propagate transversely along and radially out from the Interaction Point (IP). Closest to the IP is the VErtex LOcator (VELO) that provides high resolution tracking information in the immediately vicinity around the IP. The next stage downstream is the first of two Ring IMage CHerenkov (RICH) detectors that exploit the phenomenon of Cherenkov radiation to separate particles by their mass and provide particle identification (PID) information. The particles then pass through a strong magnetic field of 4 Tm that can operate in either a polarity up or down state, which bends changed particles through electromagnetic interactions. Theses bent paths are seen in the tracking stations (Silicon Tracker (ST) and the Tracking Turnices (TT)) where further momentum information can be extracted from the trajectories as well as providing linking information between the early and later stages of the detector. Following the tracking stations is the second RICH detector, containing a different gas mixture to RICH1 to cover a different momentum range. The two calorimeters provide information about the energy of the particle candidate, first in the ECAL for electromagnetically charged particles and photons, then in the HCAL for hadrons. The final stage is the series of muon detectors, M2 through M5 (M1 located prior to the calorimieters.) The rest of this chapter will break down each component of the LHCb detector.

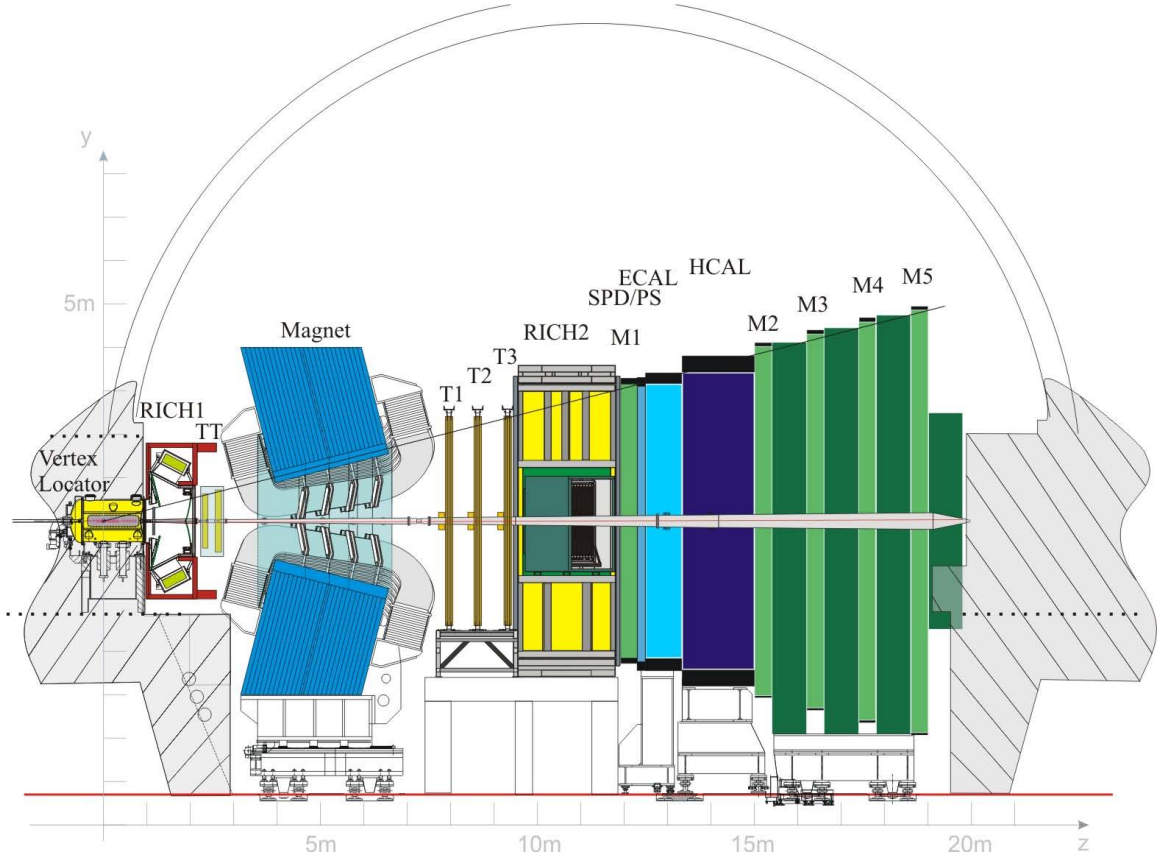

Figure 3.3: A diagram of the LHCb detector overlaid on the dimensions of the cavern.[22]

3.3 Detector

3.3.1 Interface to the LHC

To minimise material interactions, the first 12 m of the beam pipe (closest to the interaction point, in the forwards direction) inside LHCb is made of beryllium. The choice of material makes it very difficult to work with as it is highly toxic, and very fragile, but contributes very little to the material budget of the detector. The detector must be dismantled from below the beam pipe until the pipe itself is removed during maintenance, as even dropping a screw onto the pipe would be enough to break it. The remaining 7 m of the pipe within the detector is made of stainless steel and held in a state of ultra high vacuum (UHV) at 10^{-8} - 10^{-9} bar. This is essential to stop secondary interactions. The bake out process to produce the UHV conditions is a commonly used method to achieve such low pressures, where after the initial vacuum conditions are met, the chamber is heated under pressure to an extent such that volatile deposits on surfaces inside the chamber are liberated and can be extracted. The purpose of this is to remove any material that could be released over time in the hostile conditions during data taking would result in a higher overall pressure.

3.3.2 Magnet

The magnets at the LHCb experiment are essential to the physics program, as without them the momentum of the particles could not be measured, charges couple not be distinguished, and the efficiency of particle identification would be dramatically reduced.[23] The dipole electromagnet bends the paths of charged particles, from which the momentum can be measured, however low momentum particles are swept outside the acceptance of the detector. These low momentum particles can only be seen in the upstream trackers. In the interests of cost and structural feasibility warm magnets are used rather than superconducting magnets, and have an integrated field strength of 4 Tm over the 10 m length of the magnetic region. The asymmetric shape of the magnet, shown in Fig. 3.4,

induces a stronger field on the upstream side of the magnet, however the effect this has on the tracking performance and vulnerable physics channels was studied and found to be negligible. The RICH detectors are shielded to the specification of 2 mT, further details can be found in 3.3.4.1. The magnet itself is constructed of a 1500 tonne low-carbon steel yoke, and coils made of pure Al-99.7 conductor weighing an additional 54 tonnes. The conducting coils are hollow such that an internal 24mm channel is incorporated to facilitate essential water cooling.

It is necessary for accurate simulation samples, and momentum reconstruction, to map the magnetic field across the detector. To achieve the desired resolution the field strength must be mapped to a high precision and the peak of the B field in space to be accurate within a number of millimetres. The detector hall was mapped using a system of Hall probes; this semi autonomous system scans over the entire detector volume and reads the field strength to several parts in 10^4.

The magnet polarity can be flipped as it provides an elegant method to evaluate the systematics in experiments due to detector asymmetries.

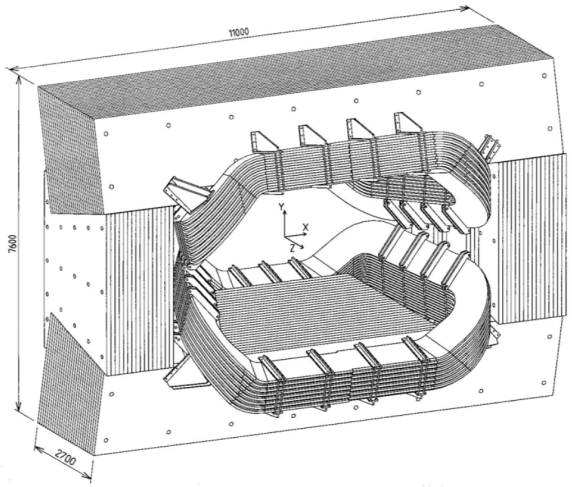

Figure 3.4: The LHCb magnet design with cross section cut out. [23]

3.3.3 Tracking and Vertexing

At the LHC, a B-meson's lifetime is approximately 1.5 ps, which corresponds to a distance in the lab of about 1 cm. Therefore, to study B-meson physics, determining the primary and secondary vertices with high precision is required.

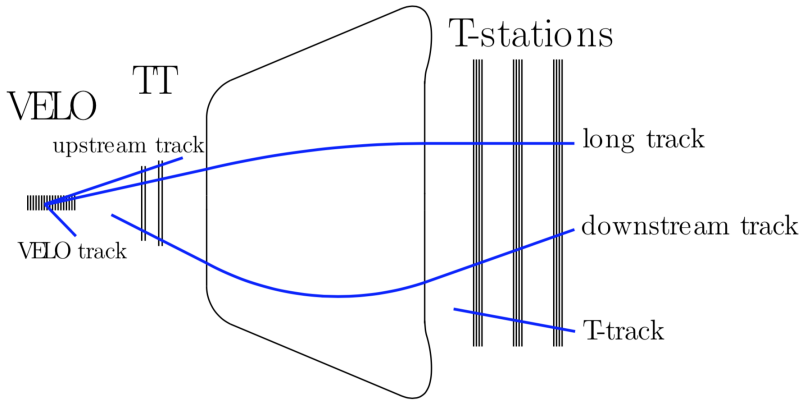

Figure 3.5: The tracking system showing five types of track upstream and downstream of the magnet. [24]

The tracking is split into three sections: the Vertex Locator (VELO); the planar Tracker Turicensis (TT) - which is upstream of the magnet; and T1-T3 - downstream of the magnet. The T1 - T3 stations are composed of two systems, the inner tracker to provide high granularity close to the beam pipe and the outer tracker to provide coverage at greater radial distances. These three tracking stations are spread over the volume of the experiment, a diagram of which is shown in Fig. 3.5, and are therefore constructed using different technologies. Each part of the tracking system can provide measurements of different types of tracks: VELO tracks and upstream tracks that are only present prior to the magnetic field; long tracks and downstream tracks that are present in detectors both up- and down-stream of the magnet; and T-tracks, only seen in the T-stations downstream of the magnet. For the VELO and TT silicon micro-strips are used, whereas for the T1-T3 a mix of micro-strips close to the detector and straw tubes further away are used.

The LHCb physics program depends on efficient tagging of B mesons, whose characteristic displaced secondary vertex require exceptional resolution in the region around the interaction point. For this purpose the VELO was developed. The VELO is a silicon

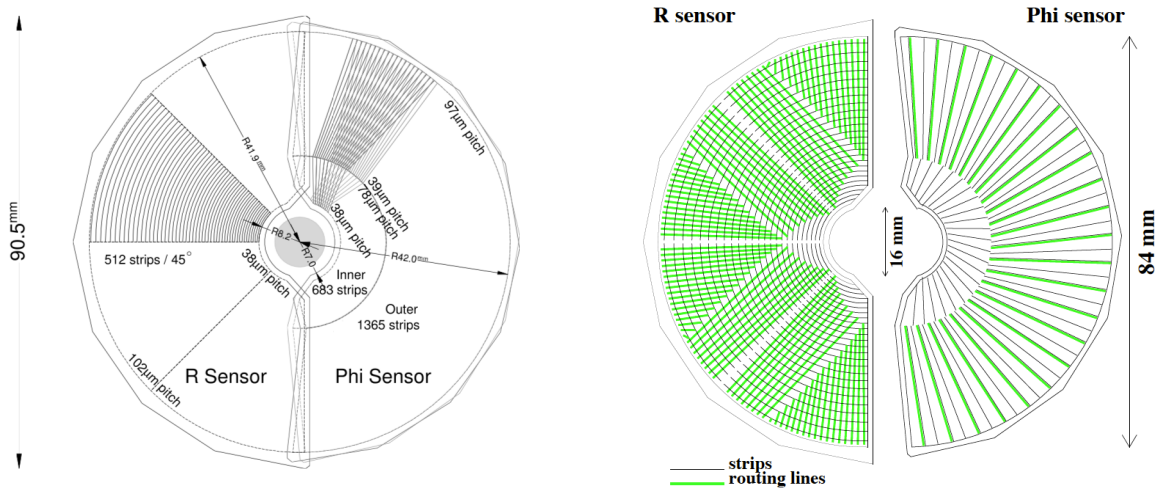

Figure 3.6: Diagram showing the VELO in the closed state during data taking (left). Diagram showing radial (R) and azimuthal (ϕ) sensor layers in the VELO (right). [25]

micro strip tracker that provides information in radial and azimuthal sensors along the beam pipe. It consists of 21 pairs of perpendicular silicon strip layers, and each layer has 2048 strips.

The VELO is a dynamic detector. During the fill period the VELO sensors are kept 6 cm apart, while during data taking conditions the radial distance from the beam to the detectors is minimised to 8 mm. This dynamic construction gives the closest measurement to the interaction point of any LHC experiment. The radial sensors are concentric arcs of silicon strips split into four regions of 45 degrees in each half of the detector, these arcs increase from 38 μm at the inner edge to 108 μm at the outer edge. The azimuthal sensors are made from a set of radial silicon strips that increase in width as distance from the beam-pipe increases. These strips are split into two sections, an inner section of 683 strips where the widths are narrower (increasing from 38 μm to 79 μm) and an outer region of 1365 strips (where the width increases from 39 μm to 97 μm). A representation of both the R and ϕ modules can be seen in Fig. 3.6. Such a complex detector is necessary to be able to increase the Primary Vertex (PV) and Impact Parameter (IP) resolution which allows suppression of the combinatorial background, where random coincident tracks from different decays are associated together. A schematic showing the displacement between

the primary and secondary vertices as a result of the lifetime of the B-meson is shown in Fig. 3.7, with PV and IP labeled. The VELO is operated in a vacuum conditions, however is separated from the beam vacuum by a thin layer of aluminium foil (known as RF foil) that prevents the harsh conditions including radio frequencies picked up from the beam from disrupting the electronics. The VELO is aligned and calibrated in real time along with the rest of the detector.

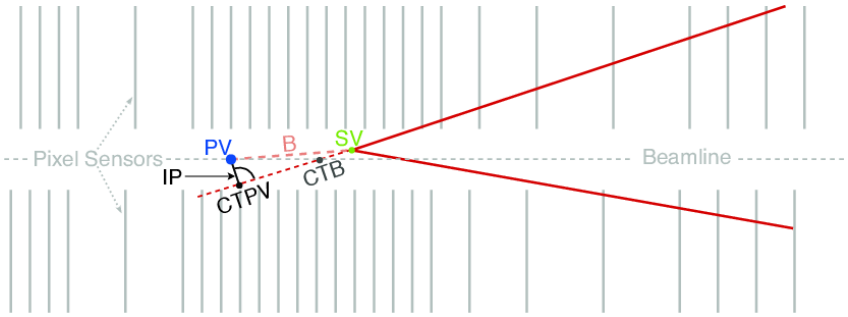

Figure 3.7: Schematic showing the displacement between the PV and SV in a B-meson decay.[26]

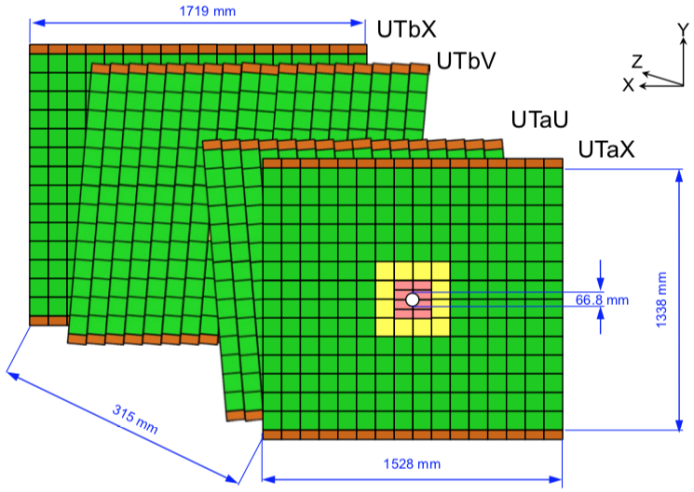

Figure 3.8: Offset layers in the ST, where the middle layers are rotated by ±5°

The silicon tracker (ST) is made of two sub detector systems: the Tracker Turicenis (TT) and the inner tracker (IT). Each module of the ST is made of four layers, where to provide coverage in two dimensions, the middle layers are offset by ±5°, as shown in Fig. 3.8. These trackers are composed of silicon micro strips, with a strip pitch of 183μm in the TT and 193 μm in the IT. The TT is located upstream of the magnet and covers

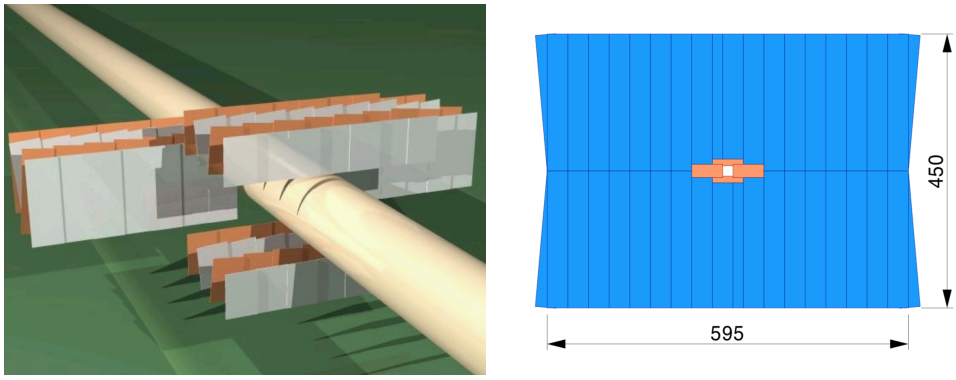

Figure 3.9: A digital rendering of an IT of a TT station (left) and a sketch of the relative size of the IT and OT (right).[27][28]

the full acceptance of the detector, while the IT is arranged in a cross formation in the middle of each tracking station downstream of the magnet, as shown in Fig. 3.9. The active coverage provided by these trackers is 8.4 m^2, read from 143360 channels, for the TT and 4.0 m^2 for the IT read from 129024 channels. The design of the ST system was governed by the following requirements:

- **Spatial Resolution** - each track must be resolved to the precision of 50 μm, as this leaves the spectrometer resolution dominated by subsequent multiple scattering effects for the full momentum spectra of the particles. Therefore in both sub-detectors a strip pitch of less than 200μm was implemented.

- **Hit occupancy** - as the radial position in the tracking system is increased the density of tracks, or hit occupancy, falls. From the inner most layer at approximately 5×10^{-2} per cm^2 to 5×10^{-4} per cm^2 in the outer layers, which means strip readout lengths must be larger in the outermost sections.

- **Single hit efficiency** - the efficiency of reconstructing hits was shown to drop dramatically for a signal to noise ratio below 10:1 from test beam studies, therefore, the design SNR is 12:1 to account for degradation over the 10 year lifetime of the detectors.

- **Radiation Damage** - in order to last the designed 10 years the ST system must be relatively radiation hard as the detectors are in the acceptance region, and in

some of the highest particle flux regions in the detector. To minimise damage from current leaking as a result of damage caused by radiation the detectors are run at an operating temperature of 5 degrees Celsius.

- **Material Budget** - as with all components in the acceptance region a concern about other sub detectors downstream is the material budget. The resolution of the LHCb detector is in general dominated by multiple scattering effects. To mitigate this the readout electrons for the TT is located outside the acceptance, and while this was not possible for the IT, the supports and electronics were kept as small and lightweight as possible.

The outer tracker (OT) is a drift time chamber, which facilitates efficient tracking of charged particles in a large volume, and being able to measure their momentum. As indication of the importance of the momentum resolution, to achieve a 10 MeV mass resolution on the decay $B_s^0 \to D_s^- \pi^+$ an error on the momentum $\delta p/p = 0.4\%$ is required. The momentum resolution is shown in Fig. 3.10. These sensors are placed in tracking stations T1, T2, T3 downstream of the magnet and cover the acceptance region of these stations not covered by the inner tracker. In the outer region of the detector the track density is lower than in the centre, so the OT can be built as an array of 'straw-tube' drift chamber modules, where each module has an array of gas tight 'straw' wires with an electrode wire in the centre, and the volume filled with a mixture of CO_2 (70%) and argon (30%) to keep the drift time low. The structure of the straw tube detector is shown in Fig. 3.11.

3.3.4 Particle Identification

Particle identification (PID) is an essential component of flavour physics experiments. At the LHCb detector the primary classifications are: pions, kaons, protons, muons and electrons. This is particularly important as these particles constitute the majority of the final products in the detector, and they present similar signals in the tracking and calorimeters. The Kaon-Pion identification is largely carried out using the RICH detectors,

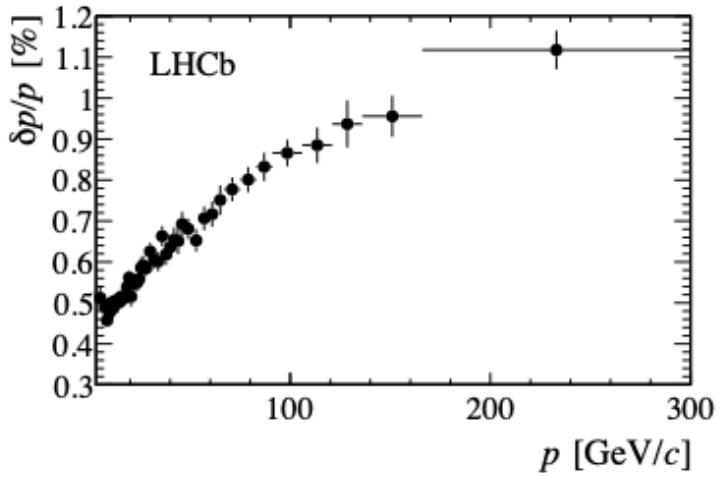

Figure 3.10: Entire detector momentum resolution as a function of momentum at the LHCb experiment, shown for long tracks from J/Ψ decays. [29]

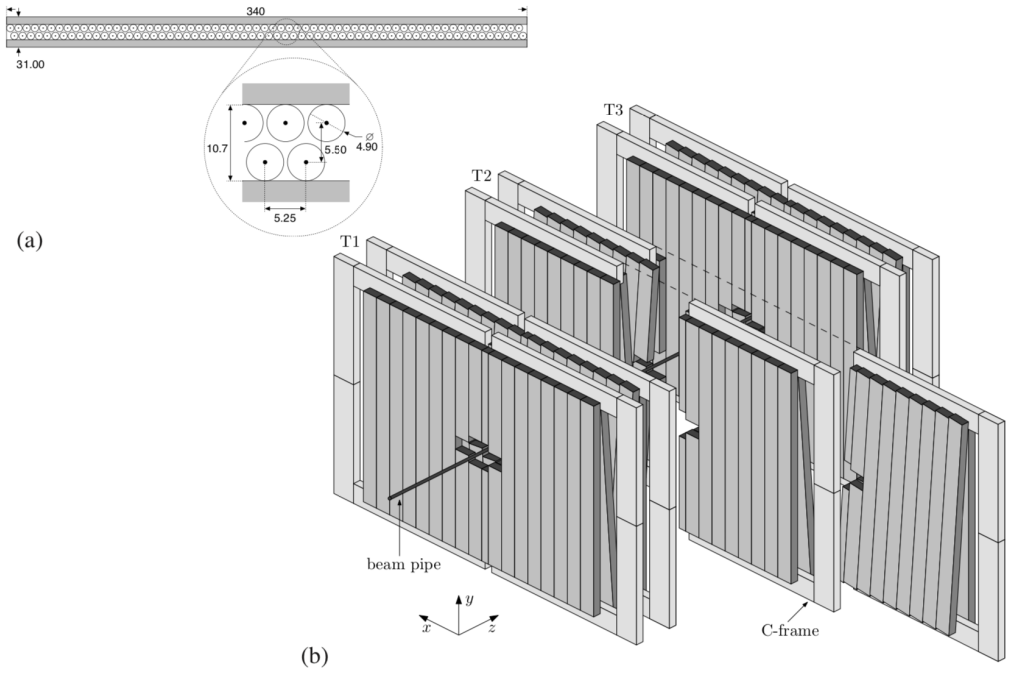

Figure 3.11: Straw Tubes and outer detector.

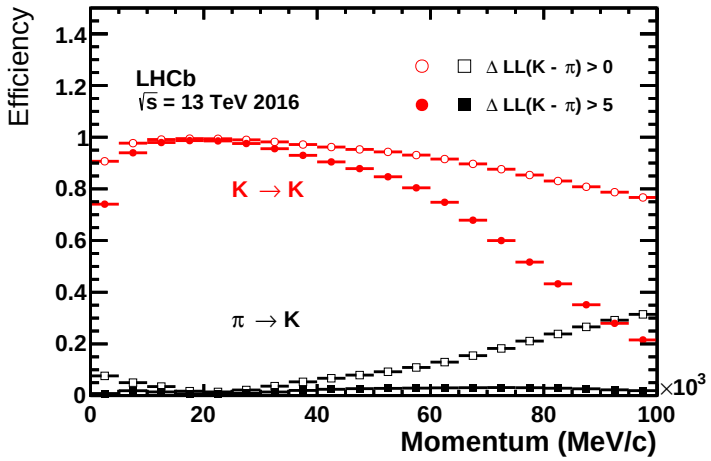

Figure 3.12: Example of PID efficiencies for Kaons and Pions as a function of particle momentum.

while the calorimenters are designed such that electrons and photons trigger the ECAL and hadrons trigger the HCAl. Muon identification is focused in the muon chambers. Additional information from the curvature of tracks in the cases of charged particles is used from the tracking stations. To collate this information a series of neural networks provide probabilities that a specific particle identification hypothesis is true. Each of the PID efficiencies is a function of particle momentum, and examples of such efficiencies are shown in Fig. 3.12.

3.3.4.1 RICH Detectors

Using the Ring Imaging Cherenkov (RICH) detectors it is possible to provide PID hypotheses for Kaons and Pions. The detectors exploit the phenomenon of Cherenkov radiation, where when a particle travels though a dielectric (or radiative) medium faster than the speed of light in that medium photons are emitted in a cone around the particle. The light cone can be characterised with an opening angle:

$$\cos\theta_C = \frac{1}{n\beta}, \tag{3.2}$$

where n is the refractive index of the medium, $\beta = v/c$ the velocity of the particle relative to c the speed of light in a vacuum. This is clearly only physically meaningful for values of $n \cdot \beta > 1$, where the opening angles varies as a function of particle velocity, and therefore

momentum of each different particle. The opening angle combined with momentum information from the tracking detectors can be used to identify different charged particles, the separation in this space of different candidates can be seen in Fig. 3.13. The RICH detectors focus Cherenkov light onto photon detectors using a series of spherical and flat mirrors, the geometry for both detectors is shown in Fig. 3.14.

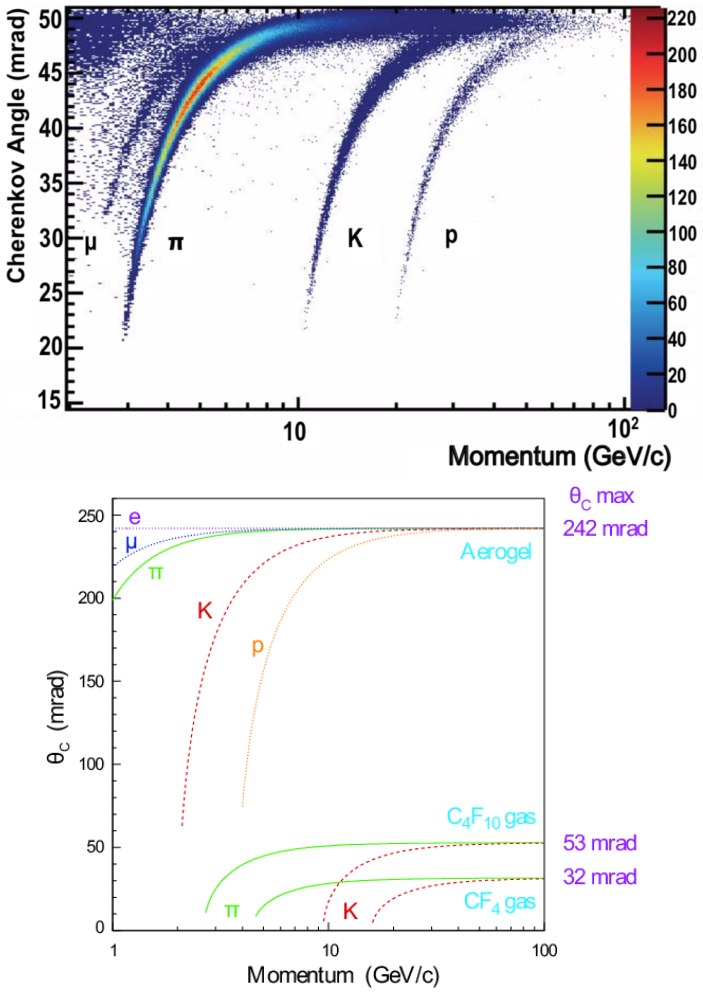

Figure 3.13: RICH θ_C opening angle as as a function of momentum provides separation of particle hypotheses, shown for both real data (top) and theoretical separation (bottom). [30]

It can also be seen from Fig. 3.13 that the momentum spectrum is harder for small opening angles of the light cone, and softer for large opening angles. Therefore, to efficiently cover the momentum spectra two separate detectors are used with different gas mixtures. The gas mixtures are chosen to tune the refractive indices so that the detectors are sensitive in different momentum ranges. The Cherenkov rings are focused onto an

array of Hybrid Photon Detectors (HPDs) outside the acceptance of the spectrometer using a set of spherical and flat mirrors. The HPDs are sensitive to wavelengths of up to 600 nm, and require a very low magnetic field, with an upper limit of 50 mTm, to operate compared to the rest of the detector. To adequately shield the HPDs a magnetic shielding system is required, which is made of iron and therefore must be outside the acceptance to avoid a huge penalty to the material budget of the detector.

3.3.4.2 RICH1

The RICH1 detector is located between the VELO and the TT, upstream of the magnet, as can be seen in Fig. 3.3. The gas medium in the chamber is flourobutane (C_4F_{10}), and during Run1 there was an additional layer of aerogel, which was removed for the Run2 data taking period. The refractive index of the RICH1 radiator covers the momentum range 1-60 GeV. In order to account for low momentum particles being swept out of the detector acceptance by the B-field, the RICH1 sub-detector is placed closer to the interaction point that RICH2. The angular acceptance of the RICH1 detector is from $\pm$25 mrad to $\pm$300 mrad, where the minimum acceptance is due to the limitation imposed by the beam pipe passing though the detector. This wide angular acceptance is motivated by similar logic to the momentum range, by catching with a wide net the low momentum particles that would otherwise be lost the reconstruction efficiency is improved. The twin factors of keeping the HPDs and their shielding outside the acceptance, and the space limitation imposed by the physical size of the cavern, means the RICH detectors are designed in a highly transverse manner. This design can be seen in Fig. 3.14 where the two sets of mirrors are used to reflect the Cherenkov radiation outside the acceptance. Spherical mirrors focus the Cherenkov cones, and the flat secondary mirrors direct the light cones onto the HPDs. The RICH1 mirror layout is shown in Fig. 3.15. As the spherical mirrors lie within the detector acceptance, they are constructed out of a carbon fibre substrate to minimise the material budget and to cope with the intense radiation in close proximity to the interaction point. The mirror construction relies on carbon fibre reinforced polymer

(CFRP) substrate, which is moulded using a polished glass mandrel, and then mounted on an array of CFRP cylinders for structural integrity. The final spherical mirrors have a spherical radius of 2700 mm, and are separated into four regions, arranged in quadrants around the beam pipe. The surface coating that provides the reflectivity is a deposition of Al (80 nm) + MGF2 (160 nm) which has the same reflectivity properties on both the CRFP substrate as with glass, so by using the carbon mirrors no performance is lost, and the radiation length dramatically improved over the alternative. The spherical mirror quadrants and cylinder mounting structure are shown in Fig. 3.16.

The secondary flat mirrors are made in a more traditional method using a glass substrate as they are outside the acceptance and do not contribute to the material budget. These glass mirrors are separated into 8 individual rectangular mirrors, and are coated with (Al + SiO_2 + HfO_2) which is the same coating used for RICH2. The impact of the optical design was studied using simulation to ensure that the optical design of the detector was not the limiting factor to the resolution performance. The photons reach the HPDs at close to normal incidence and the HPD array is made of 7 rows of 14 HPDs hexagonally tessellated.

3.3.4.3 RICH2

The RICH2 detector system operates in a very similar was to RICH1 so only the key differences will be outlined here. The second RICH detector is located between the last tracking station and the muon station, downstream of the magnet, and is tuned to be sensitive to the higher momentum range, which means a tighter opening angle of the Cherenkov cones. The RICH2 detector covers the angular range ±15 mrad to ±120 mrad in the horizontal and ±100 mrad in the vertical around the beam pipe. The sensitive momentum range is 60 GeV to 100 GeV, as low momentum particles will have been swept outside the acceptance range by the magnet after having been identified in RICH1. The radiator used in RICH 2 is a CF_4 gas. The optical system is broadly the same as RICH1 from a principle point of view, however is constructed from a larger number of mirrors.

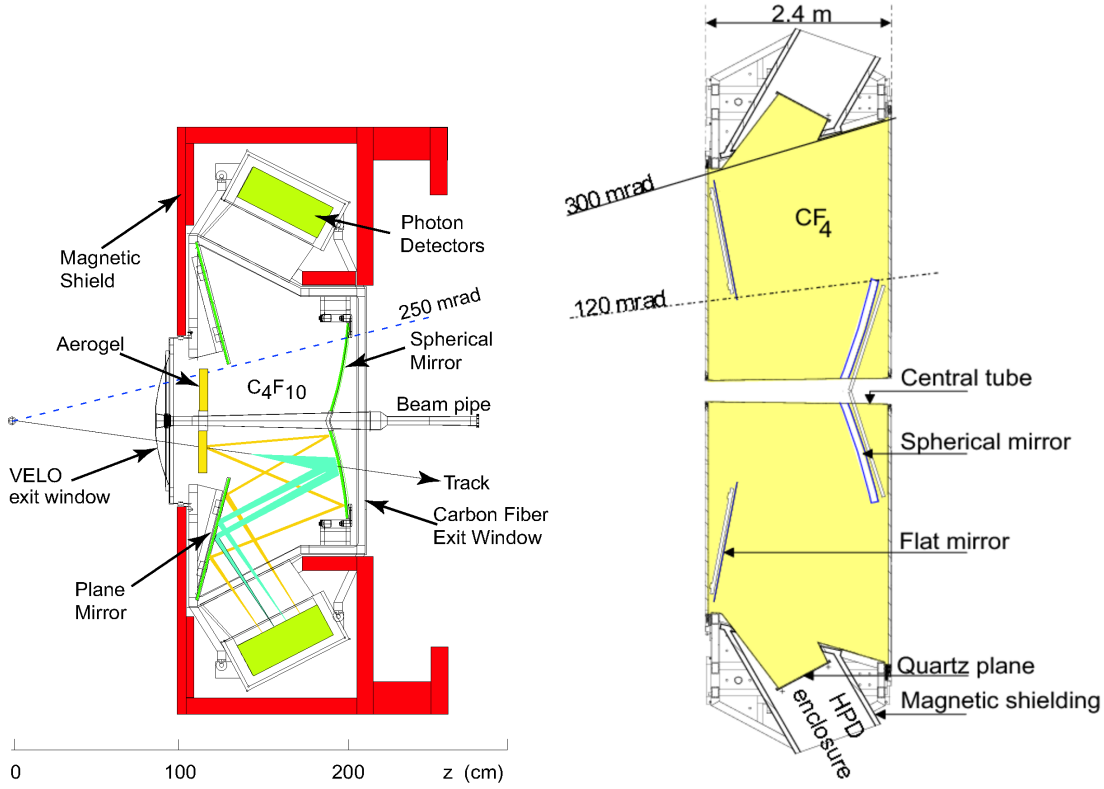

Figure 3.14: The RICH1 (left) and RICH2 (right) detectors. Note that during Run2 the aerogel was removed from RICH1. [30]

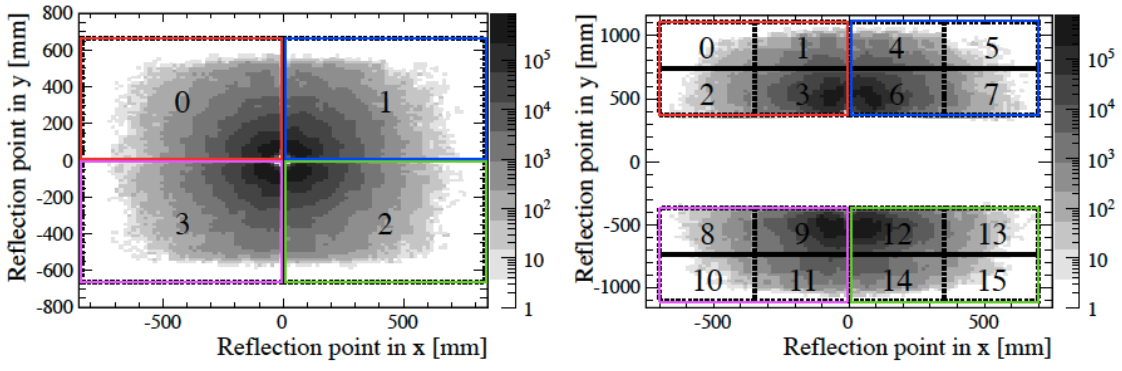

Figure 3.15: RICH1 mirror layout, primary mirrors (left) and secondary mirrors (right). The grey-scale shows the relative distribution of photon hits across the mirrors.

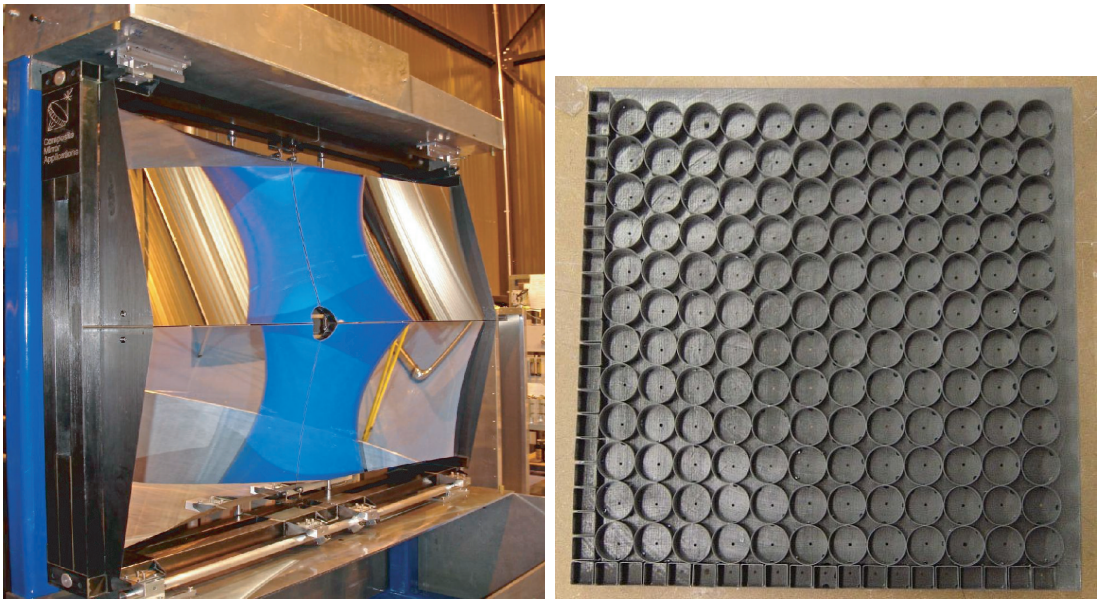

Figure 3.16: The CMA mirrors and the substructure on which they rest

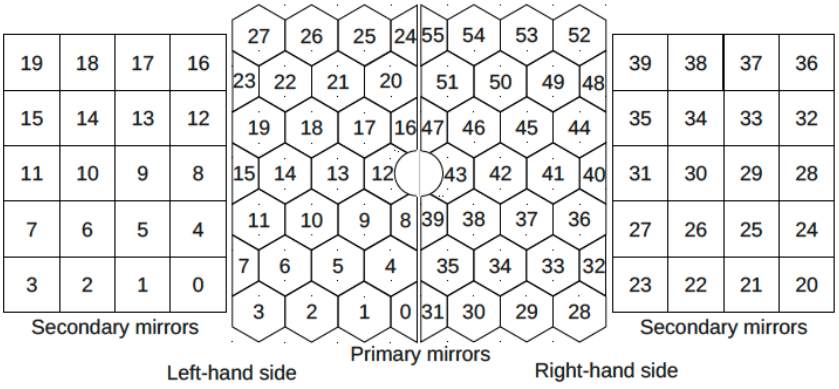

Figure 3.17: RICH2 mirror layout composed of tessellated primary hexagonal mirrors and square secondary mirrors.

The optical system is made of 26 hexagonal primary mirrors followed by 20 rectangular flat secondary mirrors in each plane of the detector above and below the beam pipe. The layout of the RICH2 mirrors is shown in Fig. 3.17. The mirror alignment system will be discussed fully in Chapter.5.

3.3.4.4 Calorimeters

The calorimeters serve three purposes in the LHCb experiment, firstly as information to the L0 hardware trigger, secondly as PID information, and thirdly as a measure of the energy of the particles. To provide a PID decision and transverse energy data to be used by the L0 trigger the calorimeter system must return a decision on an interaction in the calorimeter as photon, hadron, or electron within 4 μs of the interaction. The calorimeter system is split into the following sub detectors in sequence along the beam line: the Scintillating Pad Detector (SPD), a Pre-shower (PS), the electromagnetic calorimeter (ECAL) that detects charged particles, and the hadronic calorimeter (HCAL) that detects hadrons. The SPD and PS are built from planes of scintillating pads. The SPD and PS are separated by a distance of 56 mm where a 15 mm thick layer of lead is housed that serves to initiate the electromagetic shower. The 15 mm of lead corresponds to approximately 2.5 interaction lengths, resulting in showers caused by photons and charged leptons. Additionally, charged leptons are measured in the SPD while photons are not. The 15 mm of lead only corresponds to 0.06 interaction lengths for the hadrons, leaving the hadronic showers to be initiated at a further downstream section of the detector. The charged hadrons however also leave energy deposits in the SPD and PS like the charged leptons. From these different conditions a simple PID logic based on the calorimeter can be constructed that contributes to the PID from the RICH detectors. All of these sub detectors use a variational segmentation perpendicular to the beam axis, as the particle density decreases radially from the beam pipe.

The basic principle of the electromagentic and hadronic calorimeters is that by using an alternating series of scintillating and high density material, as an energetic particle traverses the calorimeter it causes a shower of secondary particles through interaction with the high density material. The ionising losses of these energetic particles cause excitations of the bound electrons in the scintillating material, which as the state relaxes results in radiation emitted in the form of photons. These photons are of a low energy and are collected using optical fibres and read out through an amplified photo-multiplier tube.

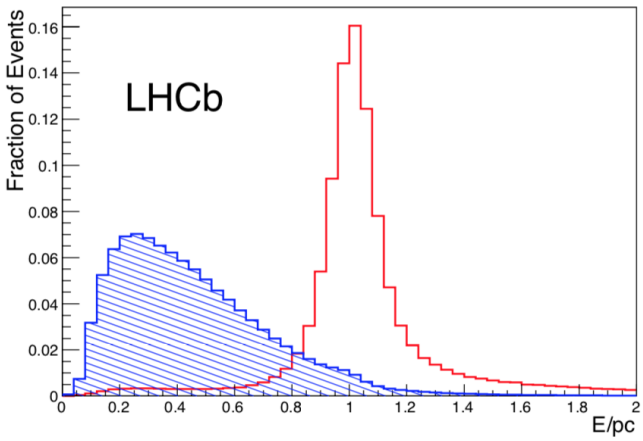

Figure 3.18: Separation of hadrons (red) from electron (blue) as a function of E/p in the ECAL.

The total energy deposit can therefore be estimated. Some particles clearly stop in the high density lead layers of the calorimieters and the energy is not measured, however these are generally the lowest energy particles and the efficiencies are well modelled.

Each module of the ECAL is built from alternating layers of 2mm thick lead, 120 μm thick reflective paper and a 4mm think scintillating tile. There are 66 layers forming a 42 cm deep ECAL block. Each scintillating layer is read out using a wave length shifting (WSF) fiber. The entire ECAL stack is wrapped in black paper to ensure optical isolation. The HCAL uses Iron and scintillating tiles, where the key difference is the scintillating tiles run in parallel to the beam pipe. This means the showers are collected along path of flight providing better lateral resolution.

There are two types of PID classifications at the LHCb experiment, using likelihood distributions, and multivariate methods. Both take as inputs a combination of information from the ECAL, HCAl and PS. Particle identity correlates strongly with the ratio E/p, an example from the ECAL for hadrons and electrons is shown in Fig. 3.18. The stochastic energy resolution of the ECAL is $1\% + 10\%/\sqrt{E}$, and the stochastic energy resolution of the HCAL is $70\%/\sqrt{E}$.

3.3.4.5 Muon System

The muon system is mostly located as the final stage in the LHCb detector, and is used to provide information on high transverse momentum (p_T) muons for the L0 trigger and PID information to the HLT in the offline analysis phase. The muon system is split into five section (M1 - M5), where M1 comes before the calorimeters, and M2-5 are downstream. Between each muon layer is an 80 cm thick slab of iron that acts as a muon filter meaning only the most energetic muons traverse all 5 stations. As can be seen in Fig. 3.19 the five stations are rectangular in shape and arranged along the beam axes, they consist of some 1380 readout channels covering an active area of 435 m^2. The geometry in the muon system is referred to as *"projective"* meaning that the spatial resolution is variable, such that the resolution in M1-M3 is superior to that of M4 and M5. The muon trigger is only passed for a high-pT muon that has five connected hits in the five layers. The purpose of the variable resolution is apparent here, where M1-M3 can provide a measure of the momentum to 20% while M4 and M5 with the lower spatial resolution are focused on identifying the highly penetrating tracks. In addition to the variable segmentation along the beam pipe there is segmentation within each plane of detectors as well. Each section above and below the beam pipe is split into 4 sections radially from the beam pipe that increase in size, and the spatial resolution in each section R1 - R4 increases following the pattern 1:2:4:8. This ensures that the flux of particles and channel occupancy are comparable between each sector.

The acceptance of the muon stations range from 16 mrad to 258 mrad. The detecting mechanism itself is multi-wire proportional chambers (MWPCs), except in M1 where the radiation damage is too high for MWPCs to last the required lifetime, so triple gas electron multipliers (GEMs) are used instead. MWPCs operate in a very similar way to the straw drift tubes discussed in Sec. 3.3.4.4, by reading a current from a cascade of electrons caused by ionisation of a gas held in a high voltage system.

In the inner most section of M1 the flux is at its highest in the muon detector, and in this region the charged particle flux can reach as high as 500 kHz/cm^2. The MWPCs

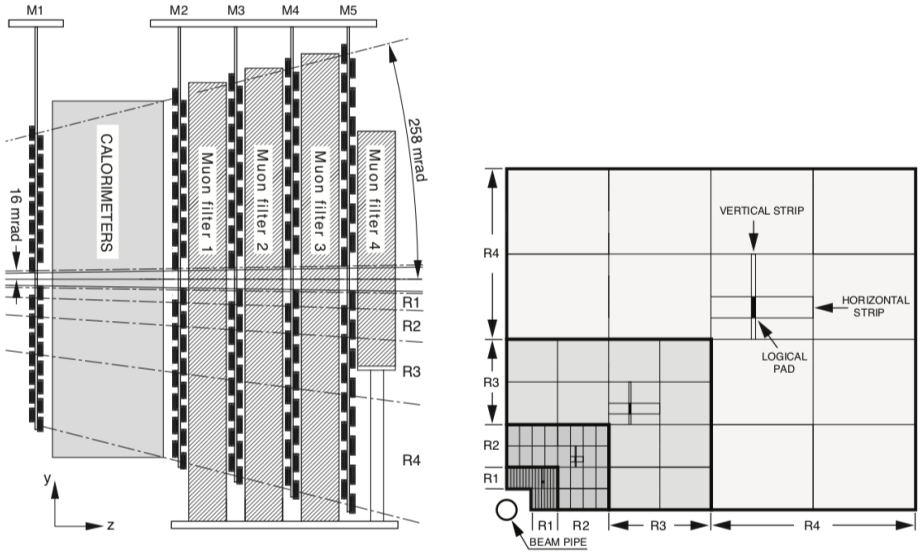

Figure 3.19: Diagram of the muon detector layers (left) and the radial dimensions of each section (right). [22]

used in stations M2 - M5 could not withstand such a flux for the operational lifetime of 10 years, so an alternative was used. The triple GEMs developed for M1 are made of two layers of superimposed such that they can operate as a logical OR. The GEM works as alternating drift gas and foil layers between an anode and cathode, so as the electrons drift from the first to the last layer they increase the gain.

As the trigger requires there to be five coincident hits in the muon stations to fire the efficiency of each must be high, and an efficiency of at least 95% is reached.

3.3.5 Trigger

The storing of data to disk at the bunch crossing rate of 40 MHz is beyond any current data storage system. In addition to the unfeasible physical write speed and storage space requirements, most of the data taken from the collisions is not useful for the physics program. A system is required to select the interesting events for further analysis: such a system is referred to as a trigger. Historically this was a task carried out by people examining individual photographic plates and looking for interesting events, however, for obvious reasons, this is now done automatically.

The trigger relies on all the sub-detectors detailed in the previous sections, and can

be generally split into three sections. The L0 trigger is at the hardware level, and looks for high transverse energy and momentum signatures. These must operate at very high rates so a decision on whether to save the event can be made before the rest of the event has been processed. This reduces the rate by collecting only events that are likely to be interesting from a physics point of view. The High Level Trigger (HLT) is split in two sections, the HLT1 performs a partial reconstruction, and the HLT2 performs the full reconstruction after another stage of data filtering.

There is a significant difference in the trigger design between the two data taking runs, in Run1 the alignment and calibration was performed offline on stored data read out from the experiment, while in Run2 the entire alignment and calibration process was moved online such that the data read out from the experiment is what is referred to as "physics ready" meaning analyses can start immediately. This data cannot be reprocessed with a different alignment and as such the alignment and calibration of the sub detectors must be stable. The first stages reduces the bunch crossing by an order of magnitude from 40 MHz to 1_MHz, by the time the second stage is reached most particles are ready to be used in classification of bunches into events. There were a number of limitations in Run1 that were specifically targeted in the redesign of the trigger system. Specifically the lack of low momentum information and only a partial PID made certain aspects of the physics program challenging, such as efficiently triggering on c hadrons.

3.3.5.1 The Run2 Trigger

In Run2 the alignment and calibration was moved *online* as part of the software HLT, this is part of a concept called Turbo stream, where the physics measurements can be performed on the output of the HLT. In Run2 the entire 40 MHz collision rate is read out, processed, and stored at 12.5 kHz in physics ready format. Flow charts showing the trigger process for Run1 and Run2 are shown in Fig. 3.20 where the addition of the alignment and calibration as part of the HLT can be seen in the 2015 trigger diagram, and the increased data rate to storage. The L0 hardware trigger looks for high transverse momentum /

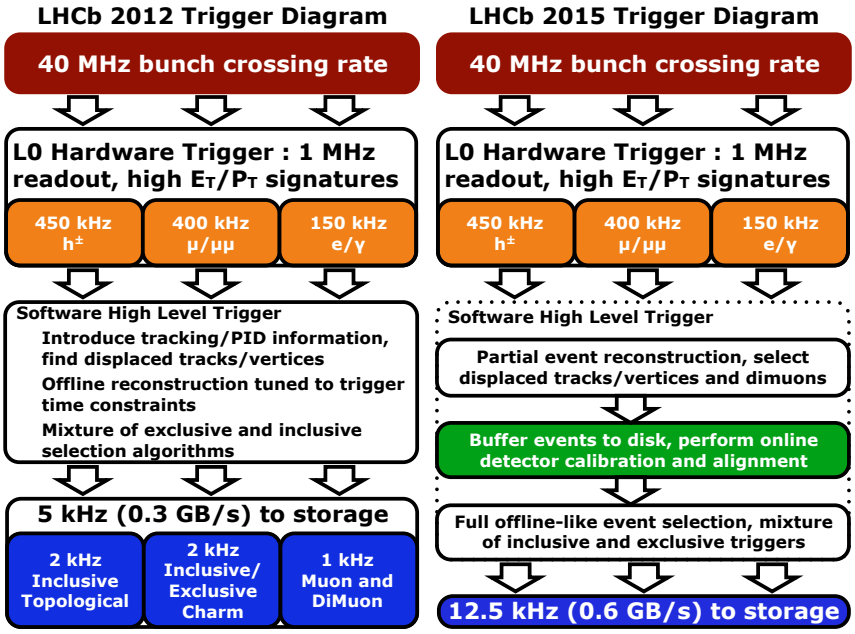

Figure 3.20: Run1 (left) and Run2 (right) trigger system comparison. [31]

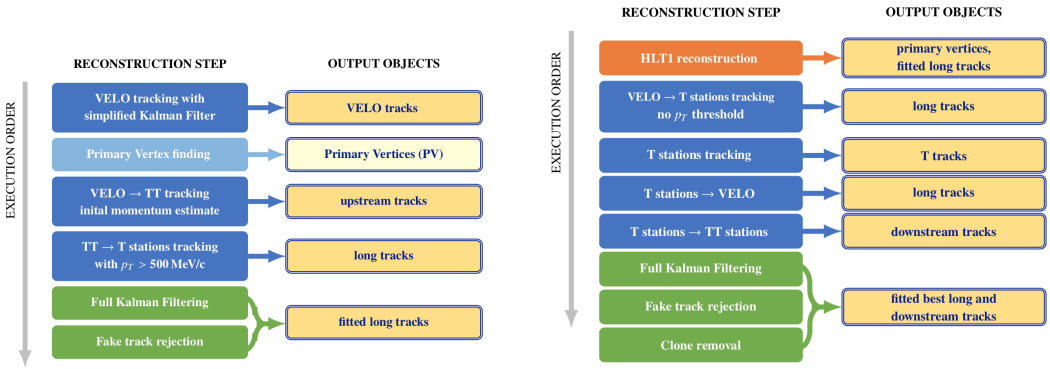

Figure 3.21: HLT1 Objects (left) and HLT2 objects (right). [31]

energy signatures, and reduces the 40 MHz by an order of magnitude to 1MHz. This data goes into the two stage software trigger, HLT1 and HLT2. HLT1 performs partial event reconstruction and reduces the rate by another order of magnitude to 110 kHz, where a 10PB buffer is used to perform the real time alignment and calibration. Once calibrated, the data goes to HLT2 where the full event reconstruction is performed with the final detector conditions applied, full PID information computed, and the final storage write rate is 12.5 kHz.

- **L0 Hardware Trigger** - the hardware decision is made based on energy deposited in the ECAL or HCAL in 2x2 clusters transverse to the beam. The energy of the

cluster is defined as $ET = \sum E_i \sin\theta_i$. Where using the logic detailed in Sec. 3.3.4.4 from the SPD and PS it is possible to identify between hadrons, photons, and electron candidates. The L0 muon trigger is slightly simpler, which is fired for straight line tracks in the 5 muon chambers. The muon track is propagated back to the impact vertex using the assumption that a single 'kick' from the magnetic field is responsible for any deviation to the path. The pT resolution is about 25% over the relevant pT range. In most trigger lines a limit is the placed on the number of SPD hits in L0 events of 450 hits in Run2 (650 for Run1) to reduce the event complexity.

- **High Level Trigger 1** - The L0 selected events are sent to the Event Filter Farm (EFF) for further processing. The sequence of reconstruction steps are shown in Fig. 3.21. Initially the VELO track are reconstructed, extrapolated into the TT stations upstream, and then the tracking stations to produce long tracks. Finally the long tracks are fitted using a Kalman Filtering method, fake tracks rejected, and the initial PV locations recomputed. The buffer plays an important role in the HLT processing, as during the 25 ps inter-bunch periods these resources can be used to provide additional processing for more complex HLT lines. Between HLT1 and HLT2 the buffer is primarily used in the alignment and calibration of the sub detectors.

- **High Level Trigger 2** - The HLT2 has a similar flow to HLT1 and can be seen in Fig. 3.21 (right). The main differences are that the inputs have already passed the HLT1, additional track types are reconstructed, and clones are removed. The tuples read out from HLT2 are physics ready, in the sense that analyses can proceed immediately with these tuples rather than requiring further stages of processing.

The EFF is composed of 1700 nodes, (800 of which are new for Run2) consisting of 27000 physical cores. It can run 50,000 single threaded processes, and runs using hyper threading. For simplicity the HLT framework is the same as that used in the offline reconstruction, which makes translation from one to the other relatively trivial. The total 10 PB buffer

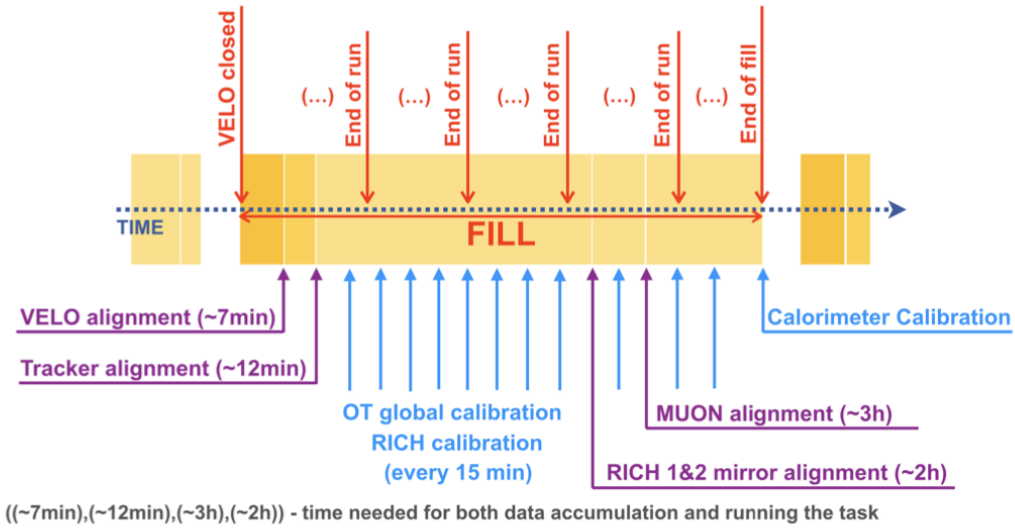

Figure 3.22: The time taken to perform the alignment and calibration of the LHCb sub-detectors during a fill. [31]

between HLT1 and HLT2 is important, as an event requires 55kB of storage post HLT1, the buffer can hold roughly 2 weeks of continuous data. The trigger lines used in the analysis presented in Chapter. 4 will be discussed in detail in Sec. 4.4.2.

3.3.6 Online System

The overhaul of the trigger for Run2 meant that data collected during the run is aligned and calibrated in real time, requiring efficient computing processes. The entire detector can be calibrated and aligned in real time in $\sim$ 4 hours after the VELO has been closed at the start of a fill, including gathering the necessary data to perform the tasks, as shown on a timeline in Fig. 3.22. In the context of the online system; alignment processes refer to spatial orientation and correction within the geometry of the LHCb detector; while calibration refers to modelling the current performance of the sub-detectors.

Initially the VELO and Tracker are aligned in 7 minutes and 12 minutes respectively, due to the relative density of data in these sub-systems they can be aligned most quickly.

The OT and RICH are calibrated every 15 minutes, however the full alignment of the RICH detectors takes approximately two hours to adequately populate all regions of the detectors. Further details of the RICH specific alignment is given in Chapter. 5. The muon alignment takes the longest of all sub-systems at approximately three hours being furthest from the PV and having the largest radial coverage in LHCb. Finally the calorimeter calibration is performed at the end of the fill. Each of these has a dedicated trigger line in the HLT1, which collects the events, and once enough are gathered the calibration is run in parallel with all other requirements of the EFF.

3.3.6.1 Stripping

Decay chains of interest are processed centrally for the LHCb experiment. Final states passing the trigger have unique selection cuts applied, and the reduced datasets are stored in a summary format. This process is referred to as stripping and allows any experimentalist to access the centrally stored data. Most analyses have a unique set of stripping criteria, the specific selection required in the analysis presented in Chapter 4 are given in Sec. 4.4.3.1.

3.3.7 Bremsstrahlung Recovery

Bremsstrahlung, or *"braking radiation"*, is a type of electromagnetic radiation emitted by a charged particle as it is deflected by another charged particle. In the context of a particle detector this is most prominent between light charged particles from the event such as electrons interacting with the atomic nucleus of the detector material.

The electromagnetic fields associated with each nucleus provide a very large number of potential interactions, and therefore a greater chance of emitting very hard photons where a significant energy loss impacts the measured properties of the electron. The energy that a charged particles looses due to acceleration in an electromagnetic field is inversely proportional to the mass of the particle squared. As electrons are by far the lightest charged particles at 0.511 MeV they suffer the greatest energy losses due to

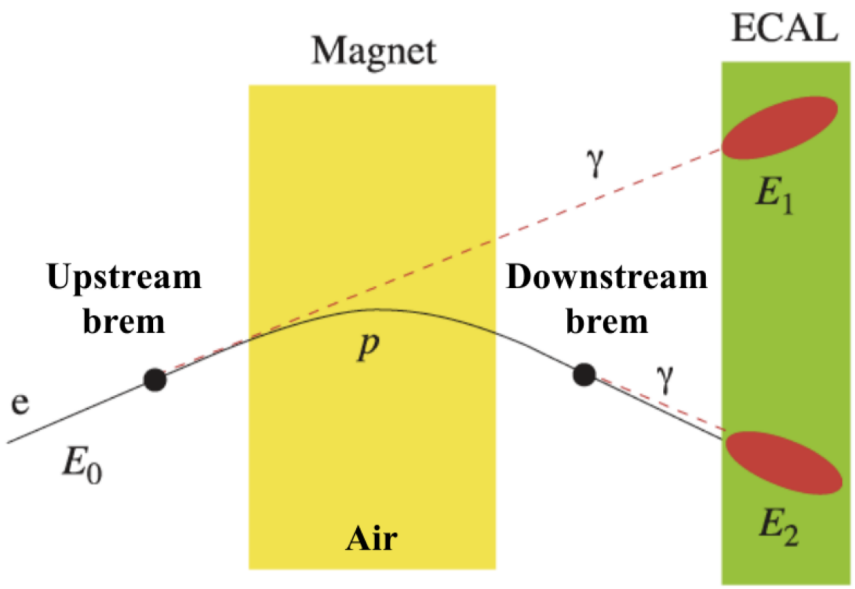

Figure 3.23: Cartoon of Bremsstrahlung radiation resulting in displaced energy deposits in the ECAL. Upstream photons result in a displaced energy deposit E_1, while downstream photons result in a single energy deposit E_2.

bremsstrahlung. The rest mass of the muon is 106 MeV, some 200 times heavier than the electron, and as such has bremsstrahlung energy losses suppressed by a factor of 40,000 compared to the electron.

Therefore, at the LCHb experiment bremsstrahlung recovery is only performed for electrons. The bremsstrahlung recovery falls into two categories as shown in the cartoon in Fig. 3.23. When a photon is emitted after the magnet a single energy deposit (E_2) is seen in the ECAL, however when a photon is emitted upstream of the magnet it results in an addition energy deposit (E_1) in the ECAL. Neutral photon candidates that are compatible with extrapolations from the electron path are added to the electron momentum to improve the resolution. In this analysis the bremsstrahlung recovery is assigned a category of either no photons recovered, one photon recovered, or two photons recovered. This obviously greatly simplifies the picture, however as soft photons contribute a small fraction of energy loss compared to hard photons this classification does not affect the physics outcomes.

Chapter 4

Angular Analysis of $B^+ \to K^+e^+e^-$ Decays

In this chapter an angular analysis of exclusive $B^+ \to K^+e^+e^-$ decays, in a low q^2 bin of $1.1 < q^2 < 6$ GeV, using data gathered at the LHCb experiment is presented. A detailed description of each section is given in Sec. 4.1. The final result is given in Sec. 4.11 and is found to be compatible with the Standard Model within a 1σ confidence interval.

4.1 Introduction

The analysis presented in this chapter uses a dataset collected by the LHCb collaboration in 2011 (at centre of mass energy $\sqrt{s} = 7$ TeV), 2012 ($\sqrt{s} = 8$ TeV), 2015 and 2016 ($\sqrt{s} = 13$ TeV). This chapter is organised into the following sections: the full breakdown of analysis strategy in Sec. 4.2, the datasets used in the analysis are described in Sec. 4.3; the initial selections are given in Sec. 4.4, the corrections to simulation in Sec. 4.5; the selection against combinatorial, partially reconstructed and semi-leptonic cascading backgrounds and the final offline selection given in Sec. 4.6; the fitting procedure to the reconstructed mass distributions and subsequent background subtraction in Sec. 4.7; the unfolding from reconstructed to true q^2 strategy in Sec. 4.8; the acceptance corrections using Legendre

moments in Sec. 4.9; the final binned maximum likelihood angular fit procedure and results are detailed in Sec. 4.10; and the systematic uncertainties of the rare mode, including from simulation and from the background subtraction method in Sec. 4.12 in addition to a number of cross checks. This analysis has close links to the published lepton universality test known as R_K[4], these links are explicitly detailed in Sec. 4.2.1.

4.2 Analysis Strategy

The following section highlights the key components of the analysis. For further in-depth details, refer to the relevant sections later in this chapter. The key aspects of the analysis strategy can be broken down as follows:

- **Control mode and rare mode** - while the angular analysis presented in this chapter is conducted using the rare decays of $B^+ \to K^+e^+e^-$, a control mode dataset of $B^+ \to J/\Psi(\to e^+e^-)K^+$ is also studied. The final state of these decays are the same, however come from different invariant di-lepton mass squared regions. The control mode serves two purposes: firstly, to validate the fit to the angular distribution; and secondly, to obtain corrections to the simulation. A measurement is presented of the angular observables in the control mode, however the procedure used to obtain the values is not optimised for the control mode, but rather serves to show that the necessary modifications to make a measurement using the rare mode do not bias the angular distribution.

- **Control mode constrained mass** - Throughout this analysis the momenta of the electrons in $B^+ \to J/\Psi(\to e^+e^-)K^+$ decays are recomputed using an engineered variable, termed the HOP mass, that exploits the known value of the J/Ψ mass, and allows the majority of background events to be removed to obtain a clean data sample. Details of this variable are given in Sec. 4.6.3. The resulting invariant mass of the 3-body $K^+e^+e^-$ system using this constraint is referred to as m_{const}. A selection is placed on m_{const} to obtain a clean sample with negligible background

contribution. When obtaining corrections to simulation a tight selection is placed as $5185 < m_{const} < 5350$ MeV, and when fitting the control mode mass distribution to obtain shape templates a looser selection is placed as $m_{const} > 5185$ MeV.

- **Corrections to the $B^+ \to J/\Psi(\to e^+e^-)K^+$ simulation** - Disagreement between simulation and data is corrected for in the control mode in two stages. The kinematic, trigger, and PID corrections derived by the R_K analysis are used. These corrections have been extensively studied in [4] and are detailed in Sec. 4.5.1. In order to ensure precise modelling of the efficiency of the angular distribution a second set of corrections are computed. These corrections are referred to as residual as they fine tune a specific set of distributions that are important in modelling the angular efficiency. These additional corrections are computed in a multivariate manner to account for correlations between distributions, and result in small differences. The $B^+ \to J/\Psi(\to e^+e^-)K^+$ is used to compute a number of corrections, however these are assumed to factorise and are done in sequence so each set of corrections relies on all previously calculated corrections.

- **Validate the corrections to the rare mode using** $\text{ProbNN}_{\text{shell}}$ - to validate that the corrections, which have been computed for the control mode, are applicable to the topology of the rare mode; a novel method using a neural network was developed. The network computes a continuous variable that ascribes a probability that the control mode events look rare mode like, this variable is $\text{ProbNN}_{\text{shell}}$. By evaluating the residual corrections in slices of this probability variable, confidence can be gained that the weighting scheme is valid. A neural network was used with the TensorFlow back-end [32].

- **Boosted Decision Tree classifier against the combinatorial background** - The combinatorial background is removed using a multivariate classifier; the Boosted Decision Tree (BDT) used is the same as in the R_K analysis for consistency.

- **Veto against the semi-leptonic cascading backgrounds** - Cascading back-

grounds are suppressed by requiring that the mass of the K^+e^- system is greater than the mass of the D^0 as $m(K^+e^-) > m(D^0)$ MeV. Cascading backgrounds where a pion is mis-identified as an electron are vetoed by requiring $m^{\text{TRACK}}_{e\to\pi}(K^+e^-) \notin m(D^0) \pm 40$ MeV, where $m^{\text{TRACK}}_{e\to\pi}(K^+e^-)$ is the mass of the Ke system with the electron reconstructed with the mass of the pion.

This veto results in a hard cut in the angular distribution above $\cos\theta_\ell > 0.65$. An attempt was made to replace the veto with a neural network classifier, to remove the cascading backgrounds and recover signal efficiency in the upper region of the angular distribution. Additional semi-leptonic backgrounds that also involved a mis-identification were discovered in the data of the type
$B \to D(\to K^*(\to K\pi_{(\to e)})X)e\nu$. These peak in the mass range $m_{K(e\to\pi)} < 1$ GeV resulting in a peak at $\cos\theta_\ell > 0.9$. Details of the original classifier are given in Appendix. A.7.

- **The analysis is performed on data split into six categories per run** - The three exclusive categories eTOS, hTOS, and TIS that describe the trigger process (full definitions of the trigger categories $eTOS$, $hTOS$ and TIS can be found in Sec. 4.4.2), and two regions of m_{HOP} split at 4900 MeV, an engineered variable described in [33] (further details in Sec. 4.6.3).

- **Fit to the reconstructed mass distribution to obtain signal and background yields** - the mass fit procedure is carried out independently for each run. Each mass fit is carried out simultaneously in the six regions described above; The separation into two regions of m_{HOP} provides a strong constraint on the part-reco yield. These yields are then used to subtract the background events from the fully selected angular distribution in the data using angular templates. The backgrounds obtained by fitting m_{Kee} cannot be subtracted from the angular distribution using an ${}_s\mathcal{W}$eight method (a process to remove backgrounds from uncorrelated distributions) due to the correlation between distributions caused by the resolution of the di-electron system[34, 35].

- **Unfolding the angular distribution to select $q^2_{\text{true}} \in [2,5]$ GeV^2 distribution** - a key point in this analysis is that the angular observables are measured to be averages in the true bin of $q^2_{\text{true}} \in [2,5]$ GeV^2, correcting for final state radiation (FSR) and detector effects. The term migrations is used to describe when an event is reconstructed in a q^2 bin that does not contain the true q^2 of the event. To that end, an unfolding method is used to correct for migrations of of decays between q^2 bins. The choice of unfolding in $q^2_{\text{true}} \in [2,5]$ GeV^2 is because the selected reconstructed region is $q^2_{\text{reco}} \in [1.1,6]$ GeV^2 Given the q^2 resolution, the bin $2-5\,\text{GeV}^2$ does not suffer from large migrations from $q^2 < 1.1\,\text{GeV}^2$ or $q^2 > 6\,\text{GeV}^2$ that would potentially introduce biases with the unfolding procedure.

- **Acceptance correction of the angular distribution** - The efficiency of reconstructing, and selecting $B^+ \to K^+e^+e^-$ decays, referred to as *acceptance*, is not uniform across the angular distribution and therefore has an impact on the final angular fit to extract the angular observables. The acceptance correction is calculated on the fully selected simulated samples with respect to the generator level distribution. The shape of the acceptance correction is modelled in segments using a Legendre polynomial and a second order polynomial. This distribution modulates the angular distribution that is used to fit the data.

- **Angular fit with uncertainties evaluated using Feldman Cousins method** - the final angular fit is performed on the combination of the three trigger categories and two data-taking periods. The angular PDF is the sum of angular distributions, each with a separate acceptance correction. Their admixture of trigger categories per run is determined from simulation, while the relative mixture of run periods is taken from the yields to the reconstructed m_{Kee} fits. The fit is a binned maximum likelihood using 20 bins in $\cos\theta_\ell$. This full fitting method is evaluated for both the control mode and the rare mode, and uncertainties are extensively investigated using toys.

  As the SM value of angular distribution parameters lies in close proximity to an

un-physical region, standard estimates of uncertainties are insufficiently robust. Therefore, a two-dimensional Feldman Cousins method is used to evaluate the statistical uncertainty of the observables of the angular distribution. The final estimates of the angular observables is given as a likelihood surface with a global best fit value.

- **Systematic Uncertainties** - Multiple sources of systematic uncertainty are considered, and their impact on the angular observables is established.

- **Cross checks** - A number of cross checks are presented in addition to the systematic uncertainties.

4.2.1 Relation to the R_K analysis

This analysis has close ties to the published 2019 R_K analysis [4], so in the interest of transparency the key overlaps are outlined in this section. Firstly the simulation and data samples are the same as in the R_K analysis. This means that in addition to the pre-processing, this analysis benefits from the extensive corrections to the simulation that correct the B-meson kinematics, the trigger efficiencies and Particle Identification (PID) response which was not done by the author. Each of these corrections was a huge undertaking in its own right, and there is a great deal of gratitude to the proponents for kindly making these samples available and providing advice at various stages of the analysis. A summary of the corrections to obtain initial weights for the simulation is given in Sec. 4.5, and a summary of the selection chain is given in Sec. 4.6. Additionally, the analysis presented in this chapter utilises the BDT developed for the R_K analysis; this ensures an identical sample post selection that facilitates cross examination of the published result. Given that these early stages of processing have been reviewed, for further details on the cross checks and validations the reader can be directed to [36]. It is hoped that this provides some additional confidence in the R_K analysis as published.

4.3 Datasets

This section describes the data and simulation samples used in the analysis, and the software stack used to process the tuples. As the analysis shares pre-processing stages with the recent R_K analysis full details of these shared steps can be found in [36].

4.3.1 Data Samples

The data in this analysis is based on $K^+e^+e^-$ combinations, that are split into two q^2 regions; $1.1 < q^2 < 6$ GeV^2 to select the rare mode region, and $6.0 < q^2 < 10.0$ GeV^2 to select the control mode region. The data used in this analysis comes from four data taking periods, 1.1 fb^{-1} from 2011 and 2.1 fb^{-1} from 2012 (referred to as Run1) and 0.3 fb^{-1} from 2015 and 1.6 fb^{-1} from 2016 (referred to as Run2). For each data taking period there is roughly an equal split of data taken with the LHCb magnetic field in a polarity down and polarity up state, referred to as MD and MU respectively.

An additional dataset of $B^+ \to K^+e^+\mu^-$ candidates is used in the optimisation and validation of the combinatorial background selection. This dataset consists of the full Run1 combinations.

4.3.2 Simulation Samples

Selection tuning and efficiency calculations are carried out using simulated events. The simulation samples are used to study the selection and geometric efficiency of the angular distribution of $B^+ \to K^+e^+e^-$ and $B^+ \to J/\Psi(e^+e^-)K^+$ decays, to model the reconstructed mass distributions of each component of the decay, and to train the neural network classifiers used in the selection against background events.

4.4 Initial Selections

The following section deals with the initial selection procedure on the tuples obtained from the stripping, the kinds of backgrounds present in this analysis, details of the kinematic

and track quality cuts, as well as the definition of q^2 and mass ranges used throughout this analysis. The stripping applies a lose requirement to track quality and particle identification criteria, as well as requiring a well defined three track vertex displaced from the primary vertex. Whenever a selection is optimised, or signal yields estimated, the branching fraction for the rare mode is taken from [37] over the whole q^2 range (including both rare and control modes) as:

- $\mathcal{B}(B^+ \to K^+e^+e^-) = (5.5 \pm 0.7) \times 10^{-7}$.

4.4.1 Types of backgrounds

There are two key types of background present in the data (for both the control mode and the rare mode) which are labelled as peaking and non-peaking backgrounds. These terms refer to the shape in the reconstructed mass distribution, m_{Kee}, of these background events, either peaking at a particular reconstructed mass (partially-reconstructed, semi-leptonic) or a smooth extended distribution (combinatorial). While this is not a complete description of all possible sources of background contribution, all others were found to be negligible. These backgrounds originate from different sources: the combinatorial stems from particles from different decays of B-hadrons being reconstructed as a single potential B candidate and passing trigger and stripping criteria. As the candidates in such a case come from different decays, the overall reconstructed mass has no peaking structure so presents as a smooth extended distribution. The shape of the combinatorial is a falling distribution as m_{Kee} increases. The primary selection against this combinatorial background is a boosted decision tree, whose details are described in Sec. 4.6.1. The partially reconstructed background comes primarily from $B^{+,0} \to K^{*+,*0}e^+e^-$ decays (or $B^{+,0} \to J/\Psi(\to e^+e^-)K^{*+,*0}$ decays in the control mode) which have a very similar description to the signal mode. Due to the enhanced width of the reconstructed mass resonances due to electron reconstruction resolution, the mass distributions of the partially reconstructed background and the signal partially overlap. Using a selection on m_{HOP}, a corrected mass for the decay accounting for bremsstrahlung losses to the electron

momentum, a portion of the partially-reconstructed background is removed. Further details and definition of m_{HOP} are given in Sec. 4.6.3 and in [33]. However, rather than simply applying a selection, the data is split in m_{HOP} and used to control the partially-reconstructed yield in a simultaneous fit in 4.10.2. In the rare mode the remaining backgrounds can be broadly classified as semi-leptonic decays. These fall into three main channels, where the branching fractions are taken from [37] as:

- $\mathcal{B}(B^+ \to \overline{D^0}(\to K^+ e\nu)e\nu) = 7.7 \times 10^{-4}$
- $\mathcal{B}(B^+ \to \overline{D^0}(\to K^+ \pi_{\pi\to e})e\nu) = 8.6 \times 10^{-4}$
- $\mathcal{B}(B^+ \to \overline{D^0}(\to K^+ e\nu)\pi_{\pi\to e}) = 16.5 \times 10^{-3}$

and are removed using a veto.

There are also peaking background where there is a single, or two mis-identified tracks. In the case of a $B^+ \to \ell\ell\pi^+$, where the single π^+ is misidentified as a K^+, these decays are suppressed by a factor of $|V_{td}/V_{ts}|^2$ and were found in the R_K analysis to be insignificant. The doubly mis-identified case where both the $\pi^\pm$ in $B^+ \to K^+\pi^+\pi^-$ are incorrectly identified as electrons has a branching fraction some 30 times larger than the rare mode so must be suppressed with particle identification selections. The R_K analysis demonstrated that these backgrounds were not significant.

At a late stage in the analysis it was discovered that decays of the type $B \to D(\to K^*(\to K\pi_{(\to e)})X)e\nu$ were a potentially significant contribution at an incidence several orders of magnitude higher than the rare mode. These decays peak in the mass range $m_{K(e\to\pi)} < 1$ GeV which results in a peak at $\cos\theta_\ell > 0.9$. These decays are removed with the same veto selection used against the above semi-leptonic decays. Details of a neural network based approach that was insufficient to remove these backgrounds completely is given in Appendix. A.7.

Table 4.1: Trigger requirements. Particle candidates indicated in parentheses means that a TOS requirement is applied on that candidate. Lines in red indicate where candidates are removed.

	Electron mode
RunI	
L0	`L0Electron`(e) $E_T(e) > 3.0$ GeV `L0Hadron`(K) & $E_T(K) > 3.5$ GeV `L0Global(TIS)`
HLT1	`Hlt1TrackAllL0`(B)
HLT2	`Hlt2Topo[2,3]BodyBBDT`(B) `Hlt2TopoE[2,3]BodyBBDT`(B)
RunII	
L0	`L0Electron`(e) & $E_T(e) > 2.7$ GeV `L0Hadron`(K) & $E_T(K) > 3.5$ GeV `L0Global(TIS)`
HLT1	`Hlt1TrackMVA`(B)
HLT2	`Hlt2Topo[2,3]BodyBBDT`(B)

4.4.2 Trigger

The efficiencies of selection vary based on how the event was selected at the L0 trigger level. Therefore, the events are grouped into two categories; events that are triggered on some component of the signal (TOS), or events triggered independently of the signal (TIS). The three exclusive L0 trigger categories are:

eTOS: these candidates require at least one of the electrons to fire the `L0Electron` trigger line. To reduce background levels a tighter offline requirement on the E_T of the electron is placed to eliminate candidates that have a transverse energy below that of the `L0Electron` threshold, but are still classified as TOS due to noise in the calorimieter. This offline requirement is $E_T > 3.0\,\text{GeV}$ in Run1 and $E_T > 2.7\,\text{GeV}$ in Run2;

hTOS: these events require that the event is firstly not triggered as eTOS, and secondly that the `L0Hadron` trigger line is fired as TOS. Similarly to eTOS, a tighter requirement is placed on the offline Kaon E_T, $E_T > 3.5\,\text{GeV}$, to reduce background levels;

TIS: These events are not triggered as eTOS or hTOS, and no extra fiducial cuts are placed other than requiring the `L0Global` line to be TIS;

- All other events are not selected by the trigger and are rejected.

The specific L0 and HLT lines are summarised in Table. 4.1, where the logical `or` of lines at the same level, and the logical `and` of L0, HLT1 and HLT2 is required.

4.4.3 The Selection Chain

The general selection chain is as follows:

1. Reconstruction and stripping selection - where the datasets are reduced to a manageable size;
2. Base offline selection;
   a) Different q^2 range selection for the rare and the control mode;
   b) Fiducial requirements to simplify the efficiency calculations, as well as track and vertex quality criteria;
   c) Particle identification (PID) requirements to reduce mis-identified backgrounds;
3. Trigger requirements;
4. Multivariate selection against combinatorial backgrounds using BDT;
5. Kinematic veto against cascading semi-leptonic backgrounds;
6. m_{HOP} cut against partially-reconstructed backgrounds used to constrain the yields in the fit to the m_{Kee} distribution;
7. Tighter mass window applied in the rare mode angular fit to further reduce background levels, $5000 < m_{Kee} < 5600$ MeV;
8. Background subtraction in $2 < q^2 < 5$ GeV range based on fit to reconstructed m_{Kee} distribution.

4.4.3.1 Stripping Selection

The stripping selection requires a loose initial track and PID quality, as well as a vertex clearly displaced from the primary vertex. The specific selections are summarised in Table 4.2.

Table 4.2: List of stripping requirements for the `Bu2LLKeeLine2` line that is used to select $B^+ \to K^+e^+e^-$decay candidates. The particle combination on which the requirement is applied is given in the first column.

			`Bu2LLKeeLine2`
B^+	$\chi^2_{to\ PV}$	$>$	100
	cos(DIRA)	$>$	0.995
	χ^2_{IP}	$<$	25
	χ^2_{DV}	$<$	9
	m (GeV)	$>$	4.28 (Run1)
		$>$	3.78 (Run2)
		$<$	6.28 (Run1)
		$<$	6.78 (Run2)
$\ell\ell$	χ^2_{FD}	$>$	16
	$\chi^2_{DV}/ndof$	$<$	9
ℓ	χ^2_{IP}	$>$	9
	p_T (MeV)	$>$	300
	PID_e	$>$	0
	χ^2_{TrackFit}	$<$	3 (Run2)
K	χ^2_{IP}	$>$	9
	p_T(MeV)	$>$	400
event	nSPD Hits	$<$	600

4.4.3.2 Initial offline selection and definition of q^2 and mass ranges

The offline selection requirements are displayed in table 4.3. The same offline requirements are applied to the signal and control modes. The q^2 ranges used to differentiate between the rare and control modes are summarised in table 4.4.

Table 4.3: Initial offline selection cuts applied to both rare and control modes.

Event quality		
nSPDHits	$<$	600 (Run1)
	$<$	450 (Run2)
ghostProb(K, e)	$<$	0.3
Cascade vetoes		
$m^{track}_{e\to\pi}(K^+e^-)$	$\notin$	$m(D^0) \pm 40$ MeV
Fiducial cuts		
$p_T(e)$	$>$	0.5 GeV
$p(e)$	$>$	3 GeV
hasRich(K, e)	$=$	`true`
hasCalo(e)	$=$	`true`
$\lvert x_{\rm ECAL}(e)\rvert$	$>$	363.6 mm
or $\lvert y_{\rm ECAL}(e)\rvert$	$>$	282.6 mm
PID cuts		
probNN$_K(K)$	$>$	0.2
pid$_e(K)$	$<$	0
pid$_e(e)$	$>$	3

Table 4.4: Rare signal and control mode samples, and their corresponding q^2 and mass ranges. These are the only requirements that are different between the rare mode and the control mode.

	ee mode
signal mode	$1.1 < q^2 < 6.0$ GeV^2
	$4.88 < m(K^+ee) < 6.50$ GeV
J/ψ mode	$6.00 < q^2 < 10.0$ GeV^2
	$4.88 < m(K^+ee) < 6.50$ GeV

4.5 Calibration of simulation to data

As part of the preparation of the tuples for the R_K analysis, extensive calibration of efficiencies and corrections of the simulated samples has been carried out. Detailed explanation of these corrections can be found in [36], but the following section 4.5.1 outlines the key stages in the procedure. Then Sec. 4.5.2 gives a detailed description of the additional multivariate corrections to the simulation. These secondary, residual, corrections ensure precise modelling of the efficiency of the angular distribution. These corrections are referred to as residual as they fine tune a specific set of distributions that are important in modelling the angular efficiency, and are computed in a multivariate manner to account for correlations between distributions. This correction procedure means that the final weights are a combination of the initial weights and the multi-variate (MV) weights as

$$w_{\text{total}} = w_{\text{initial}} \times w_{\text{MV}}. \tag{4.1}$$

This combination of weights ensures that the simulated samples are generally well calibrated across all categories, before fine tuning the kinematic variables that could bias the angular distribution.

4.5.1 Initial calibration

The total efficiency of selection ϵ_{tot} is made up from a combination of efficiencies as

$$\epsilon^{\text{tot}} = \epsilon^{\text{geom}} \cdot \epsilon^{\text{rec,strip}} \cdot \epsilon^{\text{presel}} \cdot \epsilon^{\text{PID}} \cdot \epsilon^{\text{trig}} \cdot \epsilon^{\text{MVA}} \cdot \epsilon^{\text{fitrange}}. \tag{4.2}$$

Each efficiency in Eq. 4.2 can be defined as:

- ϵ^{geom} - the angular acceptance of the tracks, where all tracks must be contained within the solid angle between 10 mrad and 400 mrad, known as the DecProDCut;
- $\epsilon^{\text{rec,strip}}$ - the efficiency of the tracking, vertex reconstruction, and stripping cuts;
- ϵ^{presel} - the efficiency of all offline pre-selection, including the specification of q^2 range to define the control and the rare mode;

- ϵ^{PID} - the efficiency of all PID cuts;
- ϵ^{trig} - the trigger efficiencies, combination of both L0 and HLT;
- ϵ^{MVA} - the efficiency of the ProbNN selections against background events;
- $\epsilon^{\text{fitrange}}$ - the efficiency of the mass window that the fit is performed over.

To get the total efficiency simulated samples are used, but first must be corrected using well understood control channels from the data to account for small degrees of mis-modelling in simulated samples. The corrections manifest in the form of three weights associated with each event. When deriving a particular correction, the correction obtained from the previous step in the correction procedure is applied. This sequence of selections is paramount to the performance of the final correction weights.

Firstly, the PID distributions are not well modelled in the simulation samples, so PID requirements are not applied. A PID is computed as the product of the PID efficiencies for each track, extracted from the PID calibration histograms: and a weight is applied on an event by event basis as $w^{\text{PID}} = \epsilon^{\text{PID}}_{K^+} \cdot \epsilon^{\text{PID}}_{e^+} \cdot \epsilon^{\text{PID}}_{e^-}$, where $\epsilon^{\text{PID}}_{K^+}$, $\epsilon^{\text{PID}}_{e^+}$, $\epsilon^{\text{PID}}_{e^-}$ are the PID efficiencies of the kaon, positron and electron respectively. Secondly, the trigger efficiencies are corrected by applying a weight accounting for the difference between data and simulation trigger efficiencies. Thirdly, the kinematic distributions are corrected to match the data. This is done in two stages, firstly, at generator level correcting the momentum and rapidity distributions, then momentum and rapidity as well as additional vertex quality variables at the reconstructed level.

What follows is a brief summary of the extensive work that can be found in [36]:

- **Truth matching** - simulated events are truth matched, a process that compares the reconstructed particle ID with the generated particle ID, and is essential to get the correct efficiency corrections for Eq. 4.2. A large fraction of simulated events are incorrectly categorised due to final state radiation (FSR), and are reconstructed as partially reconstructed background. A small fraction of other mis-identified events

are categorised as fake tracks (referred to as ghosts), where unfortunately, a large number of combinatorial events are also categorised. To suppress the combinatorial events, firstly, only events where one of the three final particles is identified as a ghost are retained. Secondly, a criterion is placed on the probability that the track is a ghost. Thirdly, applying the multivariate selection removes the remaining majority of combinatorial events. This removal ensures efficiencies are well calibrated. Signal candidates containing a large fraction of fake hits were removed.

- q^2_{true} **definition** - the methodology of this angular analysis is sensitive to the angular observables in an integrated true $< q^2 >$ region $q^2_{\text{true}} \in [2, 5]\,\text{GeV}^2$ region, where q^2_{true} is defined prior to any detector effects, and prior to final state radiation (FSR). As electrons can lose a significant fraction of energy through FSR compared to kaons and B mesons, q^2_{true} is defined as $q^2_{\text{true}} = |P^{\text{true}}_{B^+} - P^{\text{true}}_{K^+}|^2$. This was done as obtaining the true electron kinematics prior to FSR was not possible in the LHCb simulation framework.

- **PID efficiencies corrections** - as the RICH and CALO performances are not adequately described by LHCb simulations the PID information must be corrected in Monte Carlo. Calibration data samples are used to calculate PID efficiencies as a function of kinematic variables. The calibration samples used for the electrons are from the control mode $B^+ \to J/\Psi(e^+e^-)K^+$. The calibration histograms are extracted using a stripping line that does not place any PID requirement on the electrons. Further selections are placed on the kinematics and quality of track to suppress the combinatorial background, and align the selection to the rare signal mode. The tag & probe method is used, which is where a tight PID requirement is placed on the kaon and one of the electrons as the tag, and the other electron is used to probe the efficiency of a given PID cut. The tag and probe electrons are reversed and the efficiency is given as the average of the two results. The PID efficiency of the Kaon is determined using LHCb PID-calibration samples. Further details can be found in [4].

- **L0 efficiencies calibration** - to correct for imperfect trigger efficiencies in the simulation the fully selected control mode (including pre-selection and PID requirements) $B^+ \to J/\Psi(e^+e^-)K^+$ is used such that a statistically clean sample is obtained. The ratio of data to simulation trigger efficiencies is used to calculate the weight w^{trig}. These corrections rely on the tag and probe method.

- **HLT efficiencies calibration** - the HLT efficiencies cannot be extracted using an ${}_s\mathcal{W}$eight method on the data due to excessive background levels, therefore, the HLT efficiencies are computed with a "fit&count" method. These corrections also rely on the tag and probe method.

- **Kinematic corrections** - The kinematics of the B-mesons in simulation are not a perfect match to the data. Corrections to the B-meson kinematics were done in a two stage process. Initially a correction to the two dimensional $(p_T(B^+); \eta(B^+))$ histogram, then a second set of corrections to the flight distance significance of the displaced vertex relative to the PV (χ^2_{DV}), defined as the difference in χ^2 of the PV reconstructed with and without the considered B^0 candidate (χ^2_{IP})[38]. This two stage process is done four times to account for small correlations between the kinematic variables.

4.5.2 Additional Multivariate Corrections

In this analysis an additional set of corrections are applied to ensure the mis-modelling of the angular distribution of the electrons is minimised. The fine grain corrections are motivated by the arguments in Sec. 2.4 that detail the proximity of the SM prediction to the nonphysical region. A small bias in the angular distribution could push the best fit point outside these physical limits, so precise modelling of the efficiency of the angular distribution is essential. These residual multivariate corrections ensure that any remaining small bias in the kinematic distributions are minimised in the simulated samples. The additional corrections are done using the fully selected $B^+ \to J/\Psi(e^+e^-)K^+$ experimental data and simulation samples, where a selection on the mass constraint of

$5185 < m_{const} < 5350\,\text{MeV}/c^2$ ensures the data is free of background. In Sec.A.2.1.1 a study can be found demonstrating the level of the background is compatible with zero. The weights are calculated independently for each run, and each exclusive trigger category $eTOS$, $hTOS$ and TIS. The corrections derived from the control mode are applied to the rare mode, where the applicability of this procedure to the rare mode topology is studied in Sec. 4.5.3.

In standard re-weighting methods a regular problem is that when multiple distributions all need correcting they may not share the same mis-modelling despite the variables being correlated; such that by correcting one variable, disagreement is introduced in another. This can be mitigated to some extent by iteratively applying weights to a number of variables until the change to the weights come below some threshold, however, this does not completely account for correlations between variables. Another approach is to use a multivariate method, where rather than weighting one or two dimensions at a time, weights can be calculated for a larger number of dimensions at once. This has the advantage that correlations between the variables will be accounted for and the push/pull of corrections causing other disagreements is reduced. Traditionally this is not a feasible approach as it requires exponentially more events to fill a higher dimensional space adequately to re-weight in a meaningful number of bins; however, with modern data driven methods it is possible to mitigate this issue and apply weights in a fine grid in higher dimensions without sacrificing confidence in the calculated weight.

In this analysis a gradient boosted re-weighting method is used to select a multi-dimensional binning scheme to efficiently compute weights over [39]. Such a re-weighting method works in a similar way to a gradient boosted classifier, using an ensemble of regression trees to optimise a loss function over the data, where here the objective is to maximise the binned χ^2 distance. The loss function for the `GBReweighter` is different to the classifier in that it maximises rather than minimises the binned χ^2 in each leaf in the ensemble of regression trees. This re-weighting implementation can be found in the `hepml` `Python` package[40]. The variables used in training the `GBReweighter` are the transverse

momentum of the B (B^+_{pT}), the pseudo rapidity of the B (B^+_{η}), the ($B^+_{IP\chi^2}$), the quality of the secondary vertex of the B ($B^+_{EV\chi^2}$), the ($B^+_{FD\chi^2}$), and the cosine of the direction of the B (B^+_{DIRA}). These distributions are shown in Figs.[A.1 - A.6] for each trigger category and run period where the agreement between simulation and data is shown.

The weighting algorithm can be summarised as follows:

1. Build an ensemble of shallow decision trees with few but large bins to maximise the binned χ^2 distance.

2. Predict the weight for each leaf of each tree and sum over the ensemble.

3. Re-weight the input distributions with the summed weights and compare with the target distributions.

4. Repeat until the loss function is stable.

Key advantages of this tree based method over a more naive binned method are that by dynamically selecting bins to optimise the loss function there is a dramatic reduction in the number of bins required to smoothly describe the data. Additionally the bins in this high dimensional input space are selected such that the weights are applied where they are most effective. This is a higher dimensional analogy to merging bins in a standard scheme where the difference between adjacent bins is small.

Standard monitoring procedures are used in the training of the `GBReweighter` to ensure the model does not over train and the predictions can be generalised. Firstly a training and testing split is applied such that the evaluation of a model is never done on the same data it was trained on; and secondly a ten fold k-folding method is used such that the training data was split into ten equal parts, training done on nine of these parts, and the predictions made on the final tenth part[41]. The hold out part is rotated through the data so that all of the data is used to train an ensemble of ten re-weighters over the full dataset and leverage resilience to noise from the population of models. The k-folding method promotes generalisation and avoids biases arising from over-trained classifiers by

		Train		Test			
Run	Trig.	# sim	# data	# sim	# data	% train	# folds
Run1	eTOS	48071	69930	32048	46620	60%	10
Run1	hTOS	8346	10098	5564	6732	60%	10
Run1	TIS	16618	24859	11078	16572	60%	10
Run2	eTOS	138038	133055	92025	88703	60%	10
Run2	hTOS	16224	14203	10816	9468	60%	10
Run2	TIS	23739	40542	15826	40542	60%	10

Table 4.5: The number of $B^+ \to J/\Psi(\to ee)K^+$events used in the training and testing of the multivariate re-weighting algorithm. The data is shown separated by run period and trigger category, where the training data is folded 10 times, and the test data held out from training as an unbiased validation set makes up 40% of the statistics.

keeping testing and training data separate. An average of the ten models is used to obtain the final weights.

A second degree of generalisation is encouraged in the training of the re-weighting algorithm by using a hold out set in addition to the k-fold in training. The hold out set is therefore not seen at all in the training procedure and can be used as a genuine test of how well the final ensemble algorithm generalises to data as before unseen, which is vital when using the weighting algorithm calculated in the control region on the rare mode. The residual weights are calculated for each trigger category in each run separately. The statistics used in each of these stages of obtaining the residual weight corrections are shown in table 4.5.

To quantify how well the re-weighting algorithm performs on the unseen data set an additional gradient boosted classifier is trained to separate experimental data from simulation data, where only the initial weights are applied. This classifier is then tested on simulation samples with both the initial weights and the additional weights from the `GBReweighter` applied, which quantifies the improvement of how well the fully corrected simulation models the data. Such a method allows a measure of how well aligned the corrected simulation and data are over multiple dimensions. The same input variables are used as in the calculation of the weights and the same test/train/validation separation is used in the performance classifier as the re-weighting table 4.5. The performance is quantified with a Receiver Operating Characteristic (ROC) curve. The ROC curve

shows the true positive rate against the false positive rate of the classifier. In the case of evaluating the performance of the weights the closer the area under the ROC curve is to 0.5 the less able the classifier is to separate the weighted simulation from real data.

The evaluation of the performance of the re-weighter can be found in Figs. 4.1. The gain in performance for each sample is reflective of both the initial disagreement and the statistics required to robustly model the corrections. Details showing the input variables to the re-weighter and the separation classifier can be found in the Appendix A.1, split between the two runs and three trigger categories in Figs. A.1 - A.6, where the relative scale of initial mis-modelling can be seen. For brevity it should be assumed from this point forward that any mention of simulated samples have the combination of initial weights and residual multivariate weights applied as defined in Eq. 4.1.

4.5.3 Propagation to the rare mode

Corrections obtained in the control mode that are applied to the rare mode suffer from the crucial unknown; the decays are not the same, especially in q^2. By construction the corrections cannot be obtained on the rare mode, therefore to understand the viability of the corrections a proxy for the rare mode must be examined.

One solution to this problem was implemented in the R_K analysis, where a series of ratio tests were performed in concentric hyper-shells of variables that parameterise the kinematics of the decay. The detector performance (and hence the selection efficiency) depend on lab variables, such as the final state particles momenta and angles, shown in Fig. 4.2. These variables were: $\max p_\ell$, $\min p_\ell$, $\alpha_{\ell^+\ell^-}$, $\alpha_{K^+,(\ell^+,\ell^-)}$, the minimum and maximum lepton momentum,and the opening angles between the two leptons and the di-lepton system and kaon respectively, all given in the lab frame . Each hyper-shell was centred around the kinematics of the electron mode, and as a result were deliberately sub-optimal for the J/Ψ mode, but would highlight any serious systematic bias that would indicate the weights were not smoothly applied over the phase space, and therefore would bias R_K. The limitation of such a procedure lies in the large dimensionality of the

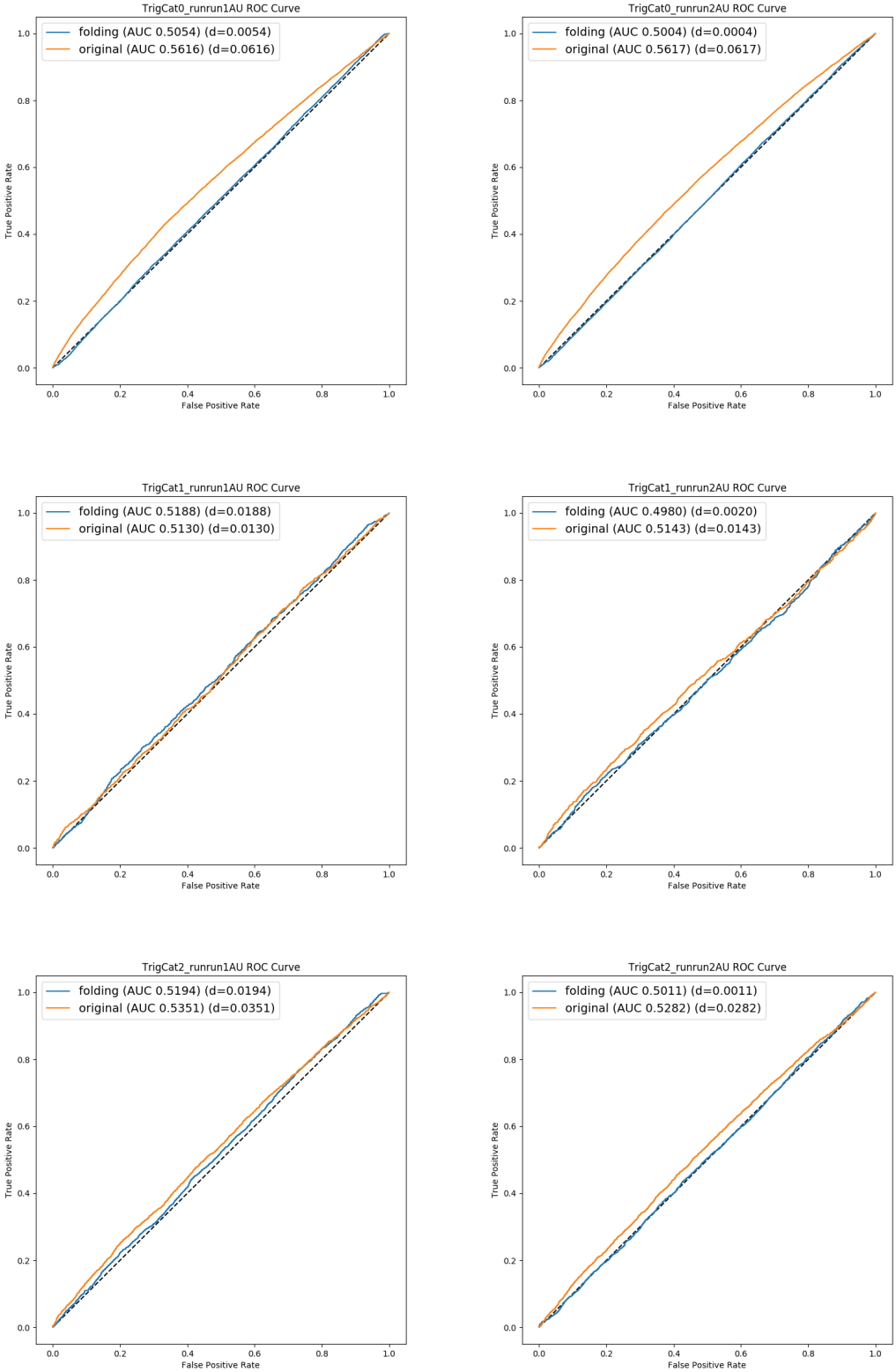

Figure 4.1: ROC curves showing the ability of a classifier to separate control mode signal simulation from control mode fully selected data with a selection on the constrained mass before and after residual correction weights are applied. Run1 (left) and Run2 (right) are split into three trigger categories eTOS (top), hTOS (centre), and TIS (bottom). The original ROC curve in orange shows the performance when only applying the initial weights to the simulated samples, and the blue shows the ROC curve for the same classifier when the residual weights are applied in addition. The key shows the area under the curve (AUC) for each curve, and the difference (d) from an AUC = 0.5 that indicates no separation.

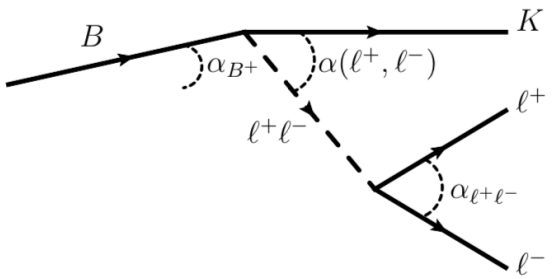

Figure 4.2: The angles and momentum in the lab frame used to parameterise the $B^+ \to K^+e^+e^-$ decays.

hyper-shells which means that only 2D projections could be tested at each time [36].

In this analysis the same concept is used but adapted to look at all four variables listed above at once. The same kinematic parameterisation of the decay is used, however these variables are instead passed as inputs to a neural network classifier that is trained to separate $B^+ \to K^+e^+e^-$ fully selected simulated samples from $B^+ \to J/\Psi(\to e^+e^-)K^+$ fully selected simulated samples, referred to henceforth as $\text{ProbNN}_{\text{shell}}$, as the output pf the network is a probability between 1 and 0 that the data is $B^+ \to K^+e^+e^-$ or $B^+ \to J/\Psi(\to e^+e^-)K^+$. To check for systematic biases that would indicate that the correction procedures outlined in Secs. 4.5.1, 4.5.2 are not applicable to the topology of the rare mode, the agreement of the angular distribution between data and simulation as a function of the $\text{ProbNN}_{\text{shell}}$ output is studied. Using $B^+ \to J/\Psi(\to e^+e^-)K^+$ simulation samples and experimental data, the angular distributions are compared in slices of the output of the $\text{ProbNN}_{\text{shell}}$ classifier, as the predicted probability places the control mode events on a spectrum from most rare mode-like to the least. This classifier can then be used as a proxy for the rare mode in the data by selecting $B^+ \to J/\Psi(\to e^+e^-)K^+$ data events that have a $\text{ProbNN}_{\text{shell}}$ value close to 0.0 as these events have a phase space description most like the rare mode.

The four variables used to train the neural network are the $\max p_\ell$, $\min p_\ell$, $\alpha_{\ell^+\ell^-}$, $\alpha_{K^+,(\ell^+,\ell^-)}$, a comparison between the distributions for $B^+ \to K^+e^+e^-$and $B^+ \to J/\Psi(ee)K^+$ can be found in Fig. 4.3. The architecture of the neural network is a fully connected dense network. There are five hidden layers, each 100 neurons wide, each

Run	decay	# train	# test	$\frac{\#sig}{\#bkg}$	% train
Run1	$B^+ \to K^+e^+e^-$	17260	4315	1.0	80%
Run1	$B^+ \to J/\Psi(e^+e^-)K^+$	17260	4315	1.0	80%
Run2	$B^+ \to K^+e^+e^-$	27169	6792	1.0	80%
Run2	$B^+ \to J/\Psi(e^+e^-)K^+$	27169	6792	1.0	80%

Table 4.6: Statistics used in the training of the $\text{ProbNN}_{\text{shell}}$ classifier. To mitigate the impact of imbalanced classes, the $B^+ \to J/\Psi(e^+e^-)K^+$ dataset is shuffled and samples are taken such that the rare mode and the control mode classes are equally populated. A train:test split of 80% is used.

hidden layer has batch normalisation applied and uses an exponential linear unit (ELU) as the activation function. The statistics used in the training of the network are summarised in Tab. 4.6, where it can be seen that the ratio of the simulation samples of each decay used in training are matched to the smaller $B^+ \to K^+e^+e^-$sample size.

The loss is calculated using the cross entropy, an initial learning rate of 0.01 is chosen and the gradient descent is optimised using the adaptive momentum method (ADAM) [42]. All input variables are normalised to a range of 0 to 1 to reduce exploding and dying gradients. In the training phase, the test and train losses, and accuracy were monitored and the model saved after each improvement in the test accuracy. The loss is monitored to ensure that this working point doesn't correspond to any significant over training.

The $\text{ProbNN}_{\text{shell}}$ distributions are shown in Fig. 4.4 for both the $B^+ \to K^+e^+e^-$and $B^+ \to J/\Psi(\to ee)K^+$simulation samples, and for both data taking runs. The separation is clear, and as expected there is a moderate overlap, which allows a continuous comparison of the angular distribution to be made using this variable.

The $\text{ProbNN}_{\text{shell}}$ variable was calculated for the $B^+ \to J/\Psi(\to e^+e^-)K^+$ fully selected data and simulation samples, and a comparison of the angular distribution was made in slices of this variable. Slices contained the same number of control-mode candidates, corresponding to 20% of the total control-mode sample. These slices can then be used to test the data/MC agreement for control-mode decays with increasing alignment to the rare mode. As can be seen in Figs. [4.5-4.6] the simulation and data remain in good agreement over the range of slices in $\text{ProbNN}_{\text{shell}}$ and only a small change to the angular distribution is observed as events are selected in a more rare mode-like phase space region.

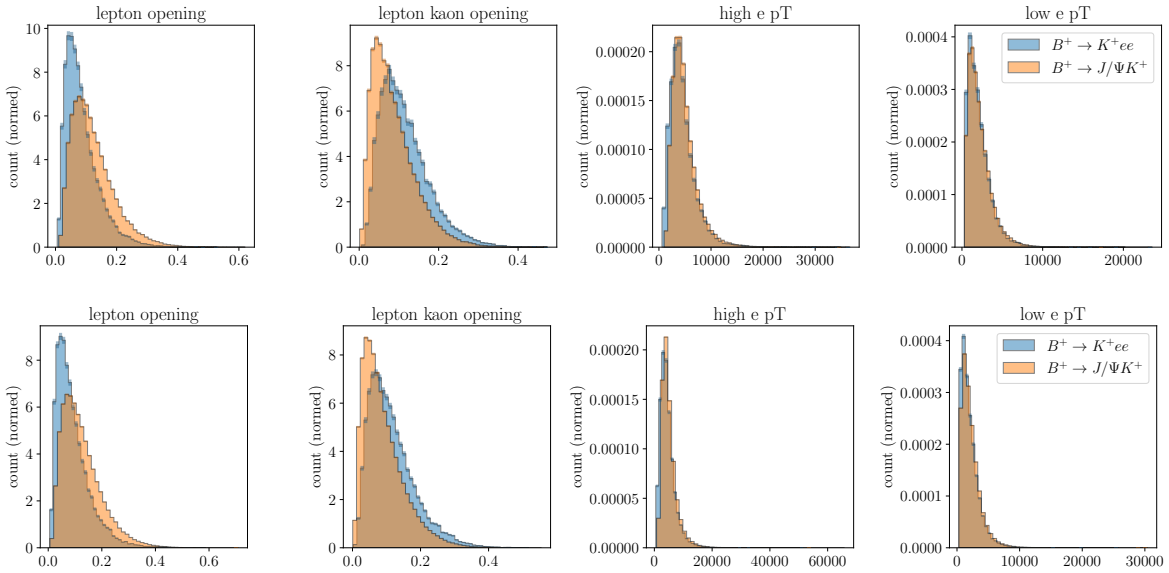

Figure 4.3: The input variables used to train $\text{ProbNN}_{\text{shell}}$ from the control mode and rare mode fully selected simulation samples. The input variables are $\alpha l^+ l^-$ (di-lepton opening angle), $\alpha_{K^+,(l^+,l^-)}$ (opening angle between di-lepton system and kaon), max pT_l (maximum lepton transverse momentum) , min pT_l (minimum lepton transverse momentum) from left to right. Run1 simulation samples (top) and Run2 samples (bottom).

The comparison is made in the three exclusive trigger categories $eTOS$, $hTOS$, and TIS for Run1 and Run2 data, and across all comparison conditions, the corrections do not seem to introduce any significant disagreement. Therefore, it is understood that the simulation models the data well even when placing a proxy selection for q^2.

The power of this method is that it becomes possible to dynamically select the events in the data where the multidimensional phase space is the closest match the phase space in the rare mode, and thus provide a best available option to validate that the corrections are phase-space independent and will not introduce bias into the rare mode. While using the $\text{ProbNN}_{\text{shell}}$ to select rare mode-like events in the control mode data is a powerful tool, it should be seen as an additional cross check rather than a replacement for the full control mode data. The full control mode samples are used to calculate corrections, and validate the mass fitting procedure, however, where appropriate the $\text{ProbNN}_{\text{shell}}$ is used to provide additional confidence that bias is not introduced into the pipeline. The resulting fits to the angular distribution of $B^+ \to J/\Psi(\to e^+e^-)K^+$ decays in the data in extreme regions of $\text{ProbNN}_{\text{shell}}$ can be found in Sec. 4.10.5.

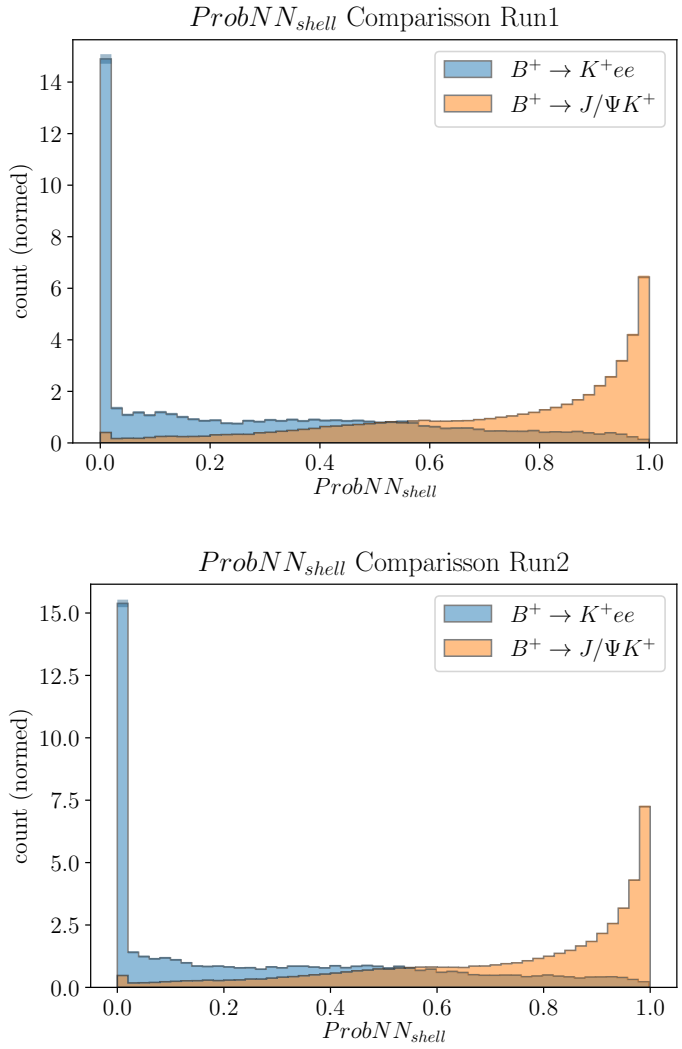

Figure 4.4: The $\text{ProbNN}_{\text{shell}}$ output on the control mode and rare mode fully selected simulation samples. Run1 simulation samples (top) and Run2 samples (bottom).

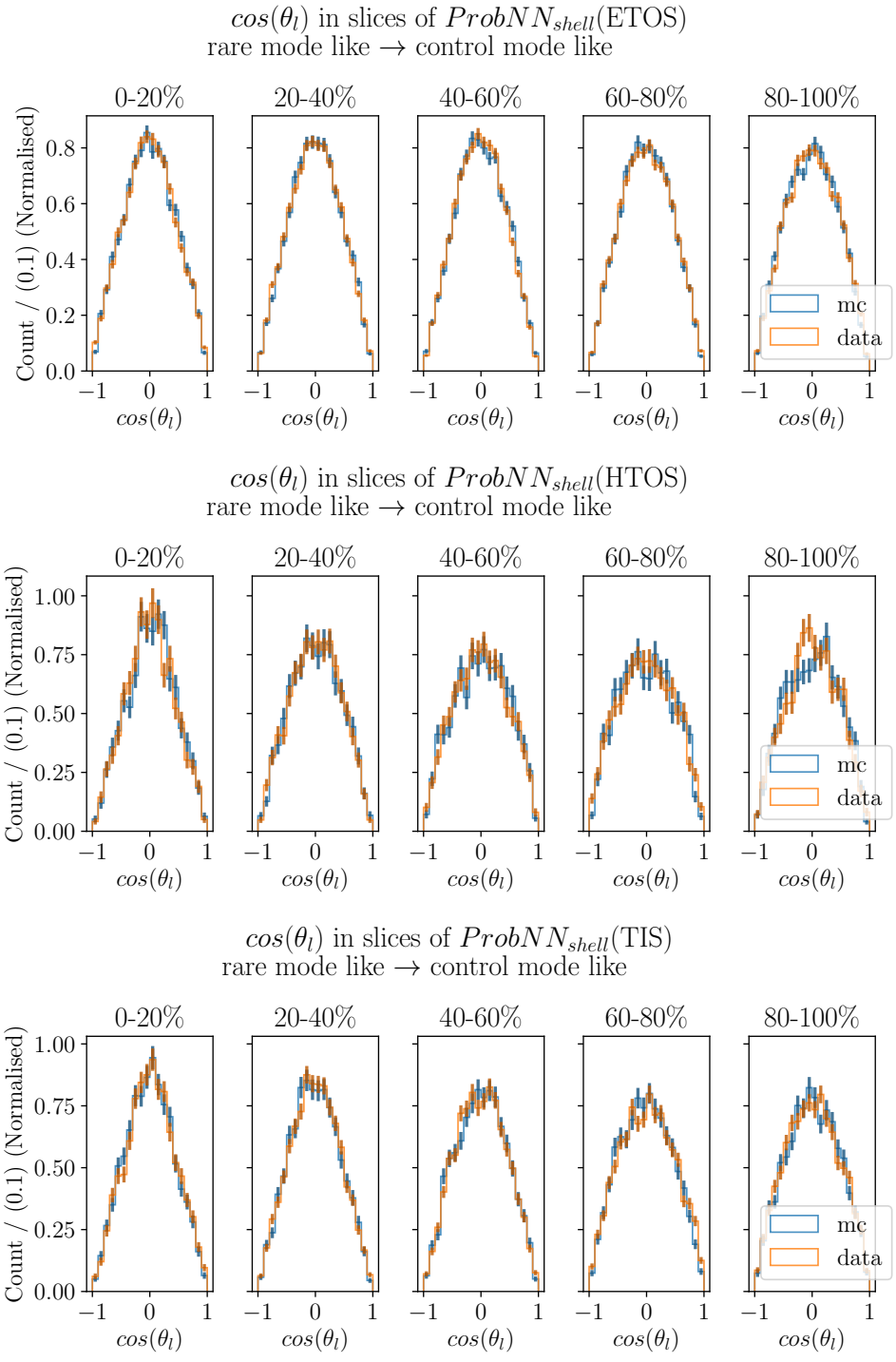

Figure 4.5: Slices in $\text{ProbNN}_{\text{shell}}$ of $cos\theta_l$, Run 1 control mode data, split into rows of $eTOS$, $hTOS$, and TIS. $0-20\%$ shows the most rare mode-like, while $80-100\%$ shows the most control mode-like.

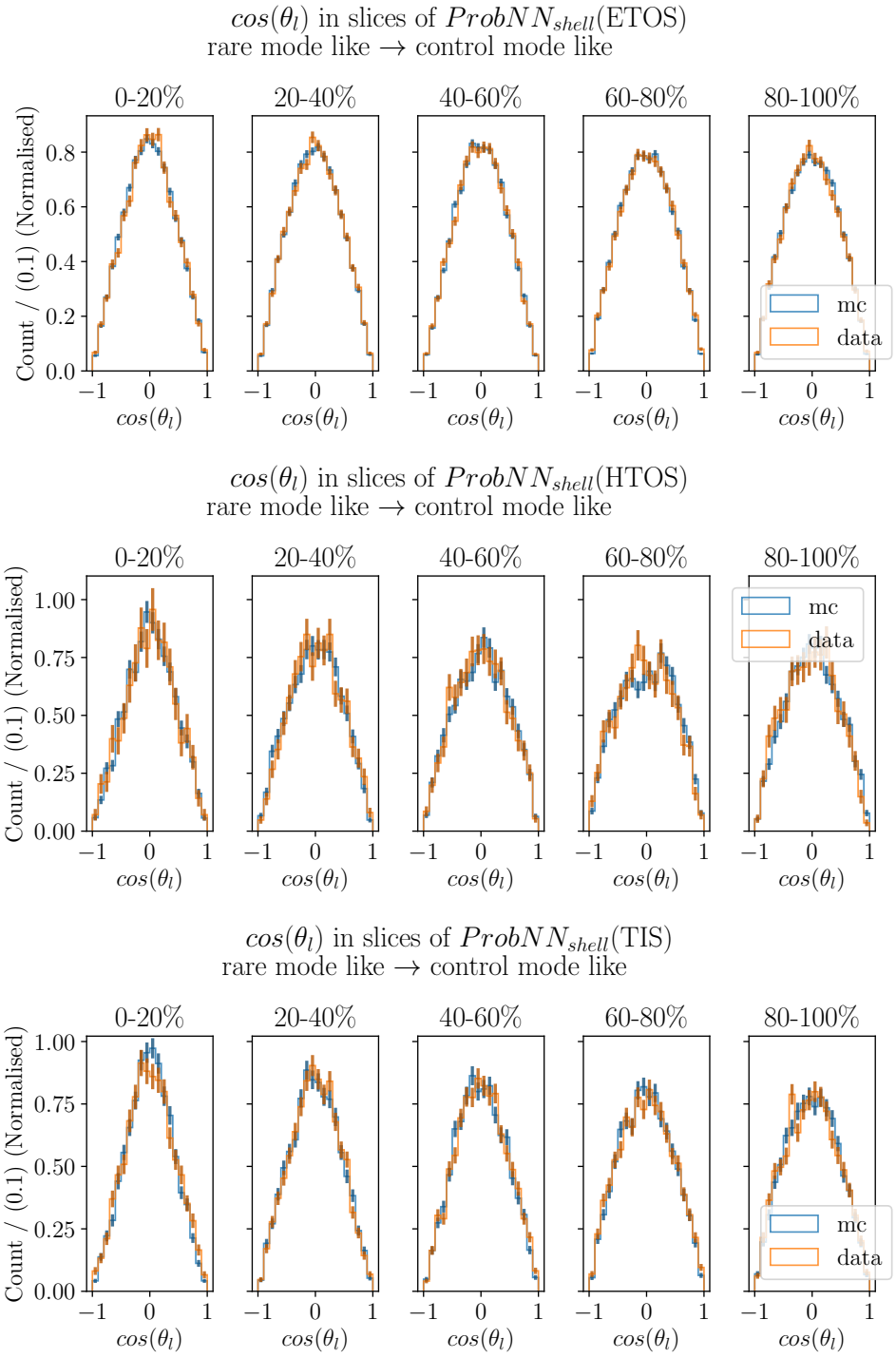

Figure 4.6: Slices in $\text{ProbNN}_{\text{shell}}$ of $cos\theta_l$, Run 2 control mode data, split into rows of $eTOS$, $hTOS$, and TIS. $0-20\%$ shows the most rare mode-like, while $80-100\%$ shows the most control mode-like.

4.5.4 Resolution Study

The resolution of $cos\theta_l$ and q^2 impact the fit procedure of the angular fit, so their resolutions were studied in the simulation to optimise the number of bins in the final fit. The difference between the reconstructed and true variable is plotted, and the resolution taken as the width of the region containing 95% of the events. Figures 4.7 and 4.8 show the difference between the true values and the reconstructed values of each event in q^2 and $cos\theta_l$ respectively, overlaid are distributions from both Run1 and Run2 for each trigger category. Vertical dashed lines represent the 2.5th and 97.5th percentiles around the median of the distributions, which is used to give a conservative estimate on how to select the width of bins in the angular fit. The distributions for both runs are in good agreement and there is little variation in the width. Similarly, the variation of the resolution in q^2 between trigger categories is small. Therefore for the unfolding procedure discussed in Sec. 4.8, a single response matrix is used for all run periods and trigger categories. Furthermore, it can be seen from Figs. 4.7 and 4.8, that the resolution impact of q^2 dominates that of $cos\theta_l$. i.e. The resolution in q^2 is a much larger fraction of the observable range than in $\cos\theta_\ell$. Only 5% of decays will result in a migration of more than $1\,\text{GeV}^2$ in q^2. Therefore, a q^2 true bin of $2-5$ GeV^2 is chosen to ensure that the unfolding technique will only need to correct for small migrations of decays between the q^2 regions. In $cos\theta_l$, a migration of 0.13 or larger will occur in 5% of the cases. A bin width of 0.1 was selected to eliminate the need for migrations between $cos\theta_l$ bins in the final unfolded data selection. However this resolution improves at the edges of the $\cos\theta_\ell$ distribution and the impact of ignoring the resolution in $\cos\theta_l$ is small, with a systematic uncertainty assigned. Other resolution combinations can be ignored due to the dominance of the q^2 resolution in the rare mode.

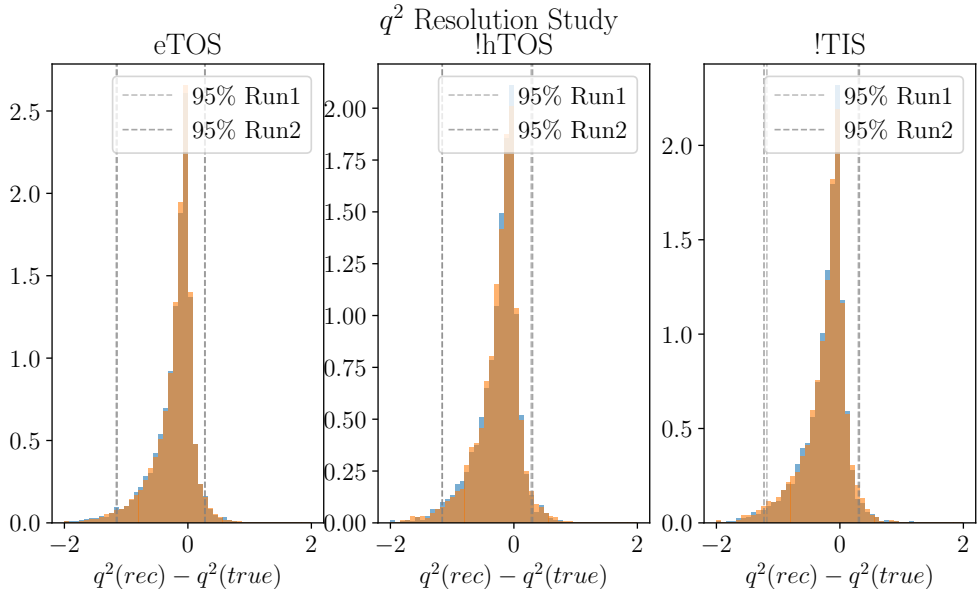

Figure 4.7: q^2 resolution study - Run1 (blue) and Run2 (orange) have identical distributions. Dashed lines bound 95% (2σ) of the events, where Run1 and Run2 have identical performances.

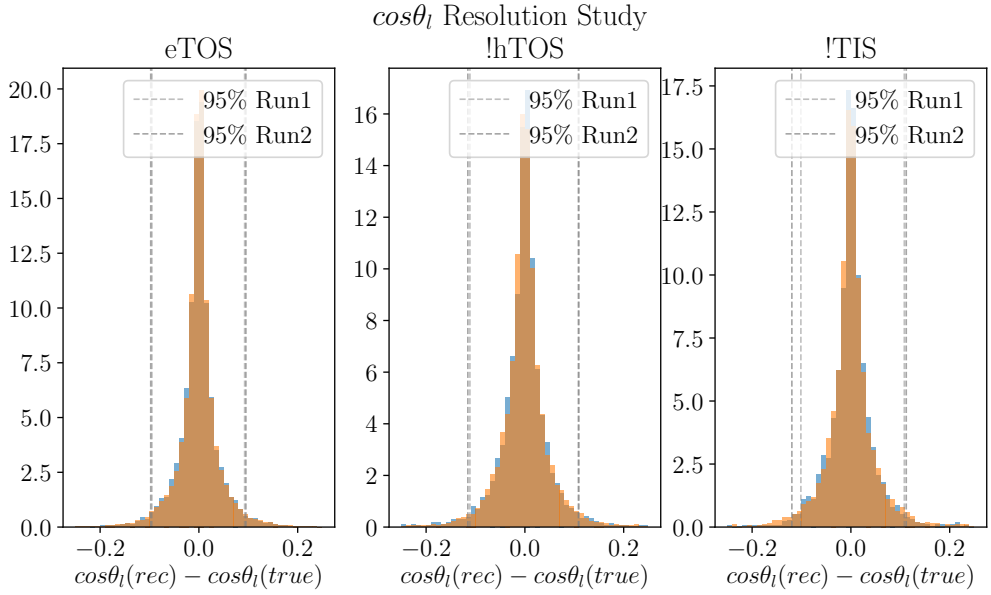

Figure 4.8: $cos\theta_l$ resolution study - Run1 (blue) and Run2 (orange) have identical distributions. Dashed lines bound 95% (2σ) of the events, where Run1 and Run2 have identical performances.

B^+	p_T, log χ^2_{IP}, χ^2_{DV}, DIRA, $\chi^2_{DV\to PV}$
$\ell\ell$	p_T, log χ^2_{IP},
K^+	p_T, log χ^2_{IP},
ℓ	min,max(p_T), min,max(log χ^2_{IP})

Table 4.7: Variables used in the BDT classifier.

4.6 Selection

The following section details the methods used to suppress the key backgrounds in this analysis, as well as studies showing the impact of the selections on the angular and m_{Kee} distributions. The combinatorial background is rejected using a boosted decision tree. The partially-reconstructed background is largely rejected using a selection on the m_{HOP} variable, after it has been used to constrain the yields in the mass fitting procedure. The cascading semi-leptonic backgrounds are rejected using a combination of kinematic selections forming a hard veto. The full selection chain is summarised in Table 4.9.

4.6.1 Combinatorial Background

This section will summarise the procedure used in the R_K analysis to train and optimise the TMVA[43] BDT. The samples used to train the BDT are the fully selected $B^+ \to K^+e^+e^-$ simulated samples as signal and $B^+ \to K^+e^+e^-$ candidates in data from the upper mass side band as background, where the upper mass side band is defined as $m_{Kee} > 5400$ MeV. All correction weights are applied to the simulated samples.

The training utilises a 10 fold k-folding method so that the entire dataset can be used for training and maximise the statistical power of the classifier. Separate classifiers are trained for each run period and in each run period a separate classifier is used for eTOS to hTOS and TIS to account for subtle differences in samples. The eleven variables used in the BDT are summarised in Table 4.7. A gradient boosting method was found to give better performance across both runs and all trigger categories than an adaptive momentum boosting method, and is therefore used to train the BDT. To chose a working point in each trigger category the significance of expected number of signal events, N_{sig},

	eTOS	hTOS	TIS
Run1	0.89	0.79	0.87
Run2	0.86	0.86	0.85

Table 4.8: BDT Working points for $B^+ \to K^+e^+e^-$decays by Run period and trigger category. Note: the eTOS BDT is different to the one used for hTOS and TIS.

against expected number of background events, N_{bkg}, is calculated as:

$$S = \frac{N_{sig}}{\sqrt{N_{sig} + N_{bkg}}}. \tag{4.3}$$

Where N_{sig} is estimated as

$$N_{sig} = \epsilon_{BDT} \cdot \frac{\epsilon^{presel}_{J/\Psi}}{\epsilon^{presel}_{rare}} \cdot \frac{\mathcal{B}(B^+ \to K^+e^+e^-)}{\mathcal{B}(B^+ \to J/\Psi(e^+e^-)K^+)} \cdot N^{presel}(B^+ \to J/\Psi(e^+e^-)K^+), \tag{4.4}$$

where ϵ_{BDT} is the BDT efficiency, $\epsilon^{presel}_{J/\Psi}$ is the pre-selection efficiency of the control mode, ϵ^{presel}_{rare} is the pre-selection efficiency of the rare mode, $\mathcal{B}(B^+ \to K^+e^+e^-)$ is the branching fraction of the rare mode, $\mathcal{B}(B^+ \to J/\Psi(e^+e^-)K^+)$ is the branching fraction of the control mode, and $N^{presel}(B^+ \to J/\Psi(e^+e^-)K^+)$ is the number of events in the pre-selected fit.

The number of combinatorial events in the signal region is estimated from the fitted yield in the upper mass side band of $K^+e^+e^-$ candidates, scaled by the expected ratio of upper mass side band events to signal region events from the $B^+ \to K^+e^+\mu^-$ sample as:

$$N_{bkg} = \epsilon_{BDT} \times N^{bkg}_{UMSB} \times \frac{N^{Ke\mu}_{sig}}{N^{Ke\mu}_{UMSB}}, \tag{4.5}$$

where the ratio $\frac{N^{Ke\mu}_{sig}}{N^{Ke\mu}_{UMSB}}$ comes from the integral of a fit to the $B^+ \to K^+e^+\mu^-$ data. The working points for each trigger category and run period are summarised in Table 4.8.

4.6.2 Selection against semi-leptonic backgrounds

The semi-leptonic backgrounds primarily arise from decays of the type $B \to D\ell\nu$ with a subsequent semileptonic or hadronic decay of the D-meson. For these transitions, one or two neutrinos are missed, or a hadron is mis-identified as an electron or kaon, resulting in rare mode-like final state. The standard approach for dealing with these backgrounds is to apply a veto against them, where any events that do not meet the criteria $m(K^+l^-) > m_{D^0}$ are rejected, or $m_{l\to\pi}(K^+l^-) \notin m(D^0) \pm 40$ MeV in the case of a pion mis-identified as an electron. The result of this cut is that the hard veto applied is particularly damaging to the electron angular distribution and results in a sharp drop off in the angular distribution for $cos\theta_l > 0.66$. This cut is significantly more damaging to the electron mode than $K\mu\mu$ modes, the poor resolution of the electrons caused by bremsstrahlung losses results in broadening of the kinematic variables that are used to construct the angular distribution. Therefore, in the electron mode a much harder, single sided, cut on $m(K^+l^-)$ is required to remove the cascading semi-leptonic backgrounds.

Such a veto naturally introduces a significant asymmetry to the angular distribution that must be modelled in order to fit the distribution. The simplest option is to apply the veto and fit up to a value in $\cos\theta_\ell$ below the drop off, this has the advantage of a simple well described acceptance correction with a single function. However, it was found in toy studies to significantly increase the uncertainty of the angular fit and introduced degenerate solutions in the form of unbound islands in the likelihood surface some distance from the expected global minimum. The approach adopted in this analysis is to use a traditional veto, and to model the drop off; this was found in toy studies to yield a mono-modal likelihood surface[1].

The angular distribution was examined without the semi-leptonic veto applied and was found to be symmetric in the control mode data and simulation, see Fig. 4.9. This symmetry can be exploited in the computation of the acceptance correction, discussed in Sec. 4.9.

[1]Details of the original classifier are given in Appendix. A.7 for posterity.

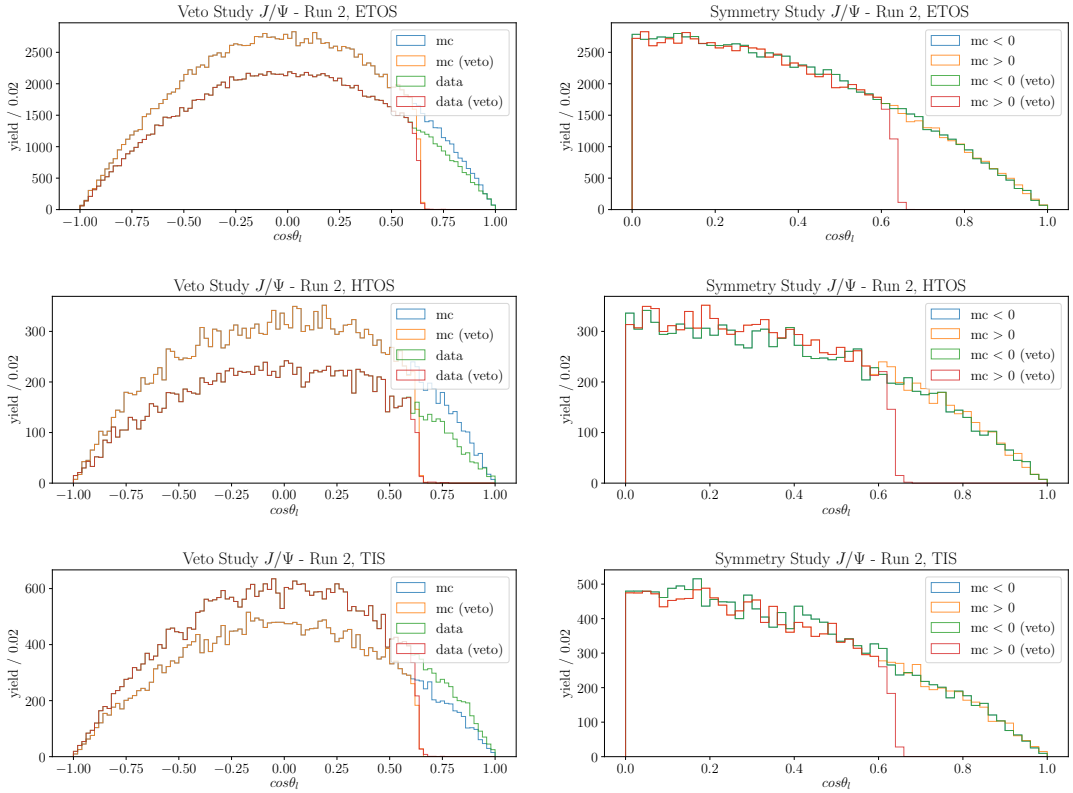

Figure 4.9: The angular distribution in eTOS (top), hTOS (middle) and TIS (bottom) for Run 2 $B^+ \to J/\Psi(e^+e^-)K^+$ simulation is symmetric in $cos\theta_l$ when the veto is not applied, and the veto can be seen as a sharp cut above 0.66. Note that these distributions are not normalised, therefore the data and simulation are not in direct comparison other than to show the drop caused by the veto is at the same value of $\cos\theta_l$. The left column shows $\cos\theta_l$, the right column shows $|\cos\theta_l|$ to demonstrate symmetry.

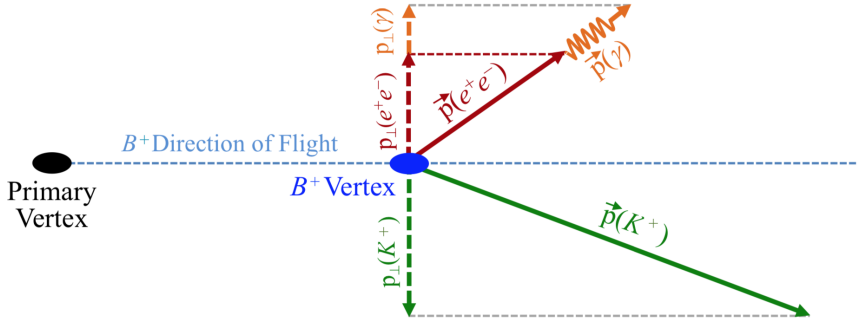

Figure 4.10: Sketch of the bremsstrahlung losses in $B^+ \to K^+e^+e^-$ decays. The bremsstrahlung losses are the difference between the transverse momentum of the kaon $p_\perp(K^+)$ and the di-electron system $p_\perp(e^+e^-)$, shown as $p_\perp(\gamma)$. Missing momentum not recovered in bremsstrahlung hits is assumed to be emitted along the line of flight of the di-electron system $\overrightarrow{p}(e^+e^-)$ as $\overrightarrow{p}(\gamma)$. Further details and original figure in [38].

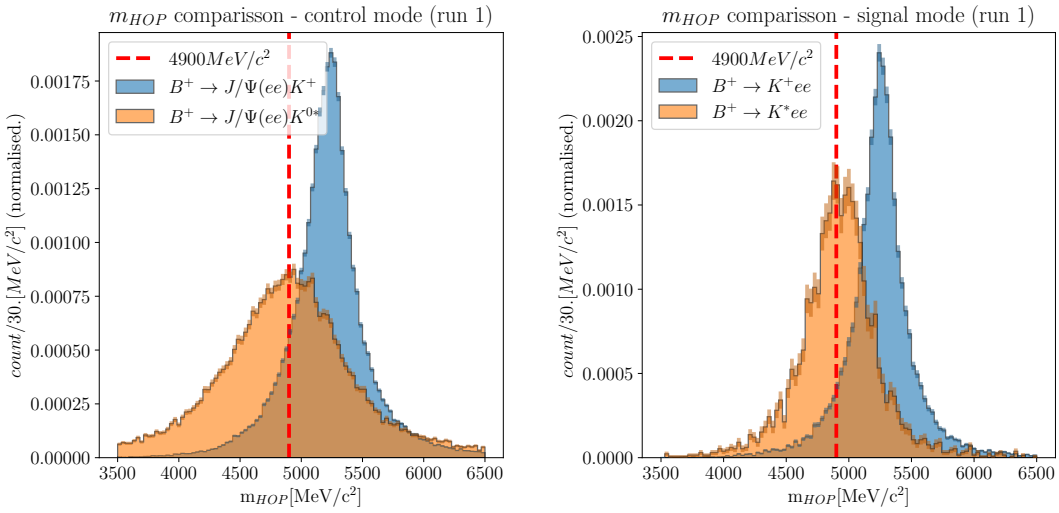

Figure 4.11: HOP mass distribution for simulated $B^+ \to J/\Psi(e^+e^-)K^+$ and $B^+ \to J/\Psi(e^+e^-)K^*$ decays (left) $B \to K^*e^+e^-$ and $B^+ \to K^+e^+e^-$ decays in Run1. (right)

4.6.3 Partially Reconstructed Background

To control the partially reconstructed background in the fit to the reconstructed m_{Kee} distribution a novel control region using the m_{HOP} variable was used. The variable m_{HOP} is defined as the corrected estimation of the reconstructed B mass that balances the momentum transverse to the line of flight of the B. This is to account for momentum losses due to bremsstrahlung that smear the resolution [33]. The magnitude of momentum of the kaon and the di-electron system transverse to the line of flight of the B-meson

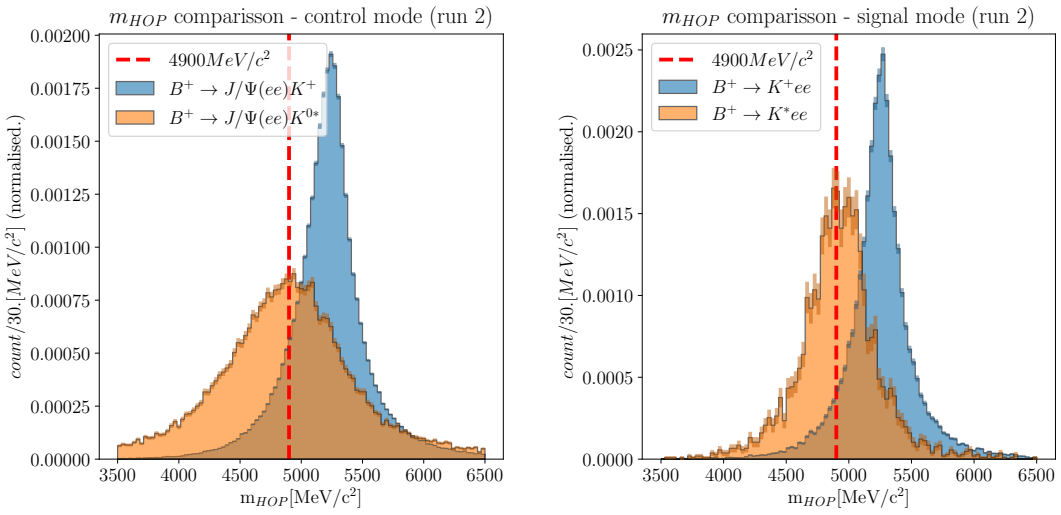

Figure 4.12: HOP mass distribution for simulated $B^+ \to J/\Psi(e^+e^-)K^+$ and $B^+ \to J/\Psi(e^+e^-)K^*$ decays (left) $B \to K^*e^+e^-$ and $B^+ \to K^+e^+e^-$ decays in Run2. (right)

are expected to be identical. Any differences are therefore assumed to be caused by bremsstrahlung energy not recovered in the di-electron system. These assumptions are shown in Fig. 4.10. The m_{HOP} is calculated by correcting the momentum of the di-electron system with the variable α_{HOP}, which for $B^+ \to K^+e^+e^-$ decays is defined as:

$$\begin{aligned} \alpha_{HOP} &= \frac{P_\perp(K^+)}{P_\perp(e^+e^-)}, \\ \overrightarrow{P^{\text{corr}}}(e^+e^-) &= \alpha_{HOP} \times \overrightarrow{P}(e^+e^-). \end{aligned} \tag{4.6}$$

The HOP mass provides a discriminating variable between signal decays and partially reconstructed background decays due to the missing π in the partially reconstructed backgrounds.

This method to constrain the background consists of separating the data into two regions, one where the partially-reconstructed background is a dominant factor (part-reco enriched region), and one where the partially reconstructed is reduced compared to the signal events (part-reco depleted / signal region). The purpose of this separation is that in the part-reco enriched region the part-reco has much greater significance against the combinatorial background and the signal events as it makes up a larger fraction of the sample, and therefore the yield can be obtained with greater precision. Using the increased precision of the partially reconstructed yield in the part-reco enriched region, it is possible

to constrain the part-reco yield in the signal region through the use of a simultaneous fit, using the relative efficiencies of the partially-reconstructed simulated samples. The comparison between background and signal distributions in the m_{HOP} discriminating variable can be seen for the control mode and the rare mode in Fig. 4.12, where the selection point at 4900 MeV/c^2 is shown. The correlation between m_{HOP} and m_{Kee} for the signal and partially reconstructed background is shown in Fig. 4.13, where a clear distinction can be seen in the distributions.

To obtain the yields for each signal and background component each run is fit separately. In each run, a six part simultaneous fit is performed to the part-reco depleted and enriched regions, across the three trigger categories. This allows two parameters to be shared in the fit, the total signal yield across all trigger categories in the $m_{HOP} > 4900\,\text{MeV}$ region, and the total partially reconstructed yield, across the three trigger categories and both m_{HOP} regions. Additional constraints are imposed by fixing the relative signal yields between the three trigger categories from simulation samples in the $m_{HOP} > 4900\,\text{MeV}$ region. The relative fractions of part-reco in each of the six regions are fixed from simulation samples, as this leverages improved sensitivity to this yield in the $m_{HOP} < 4900\,\text{MeV}$ to correctly fit the yield in the $m_{HOP} > 4900\,\text{MeV}$ region. Then for all further steps in the analysis the part-reco enriched region is discarded. This has only a small suppression on the rare mode signal yield, as can be seen in Fig. 4.17, but serves two purposes for controlling the background. Firstly, by splitting the data into the depleted and enriched regions, the precision on the yield of the part-reco yield is greatly improved. Secondly, as can be seen in Fig. 4.17 and Fig. 4.15, the background levels are reduced. Further details of how this is used to constrain the yield of the partially reconstructed background can be found in Sec. 4.10.2.

Placing a cut on the m_{HOP} variable can potentially sculpt the m_B distribution of the signal and the combinatorial background in a non-trivial way. In order to test the impact of the m_{HOP} cut on the reconstructed mass distribution of combinatorial like events a study using a sample of $B^+ \to K^+e^+\mu^-$ data passing all selections was conducted.

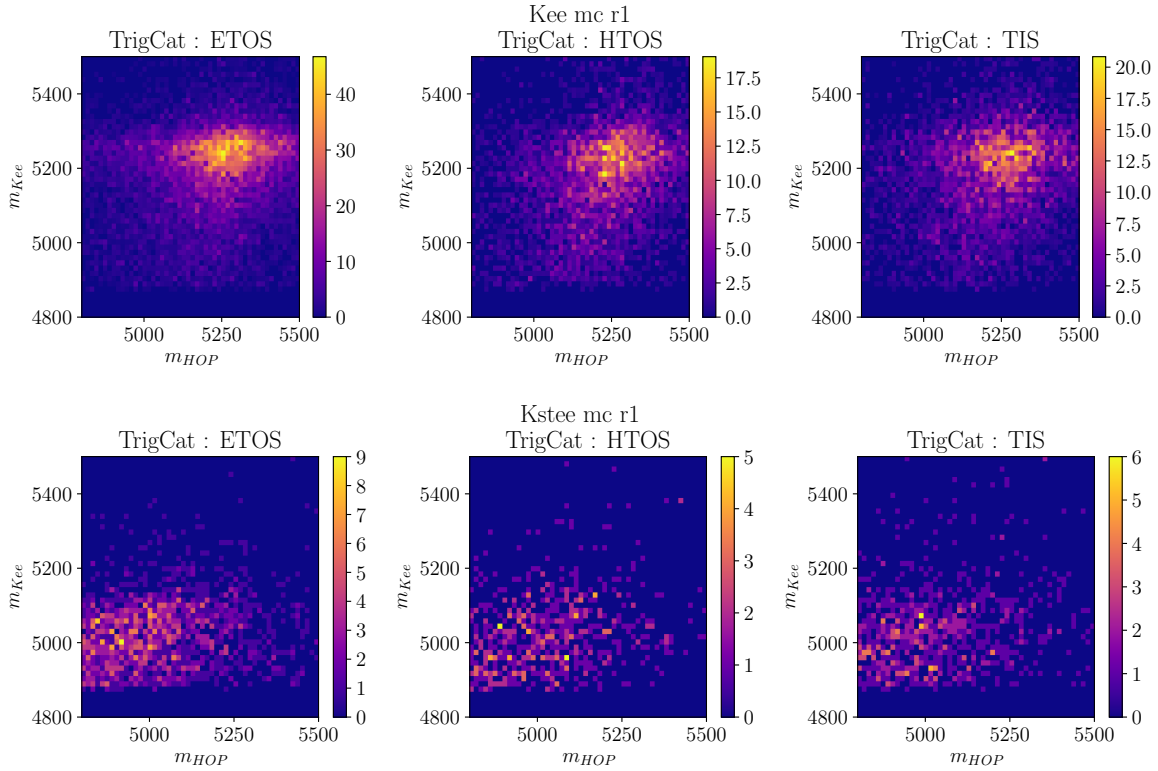

Figure 4.13: Example correlation plots showing m_{Kee} against m_{HOP} for $B^+ \to K^+e^+e^-$ simulation samples (top) and $B^{+,0} \to K^{*0,+}e^+e^-$ simulation samples (bottom) for Run1.

This sample is expected to be dominated by combinatorial events and contain no signal. Figure 4.14 shows the $m_{Ke\mu}$ distribution for events falling above and below the m_{HOP} cut. The loss of events due to the m_{HOP} cut is smooth across the whole reconstructed mass range, meaning that the exponential PDF used in the mass fit to describe the combinatorial background is still applicable. As a secondary point the efficiency of events passing the $m_{HOP} > 4900$ MeV is 74% which provides an additional method to remove combinatorial events in addition to the multi-variate selections. These stacked histograms show the efficiency of selection and rejection based on the m_{HOP} cut, the lower section of each distribution shows the portion of the distribution selected into the signal region, and the upper section shows the portion used in the low m_{HOP} fit, and then rejected.

A second check for combinatorial sculpting uses the full $B^+ \to K^+e^+e^-$ data sample, with no multivariate combinatorial selection, so that the background dominates. The study was made here with the signal region blinded. The result of this study in Fig. 4.15 shows the combinatorial is reduced significantly by this selection. While the cut is not

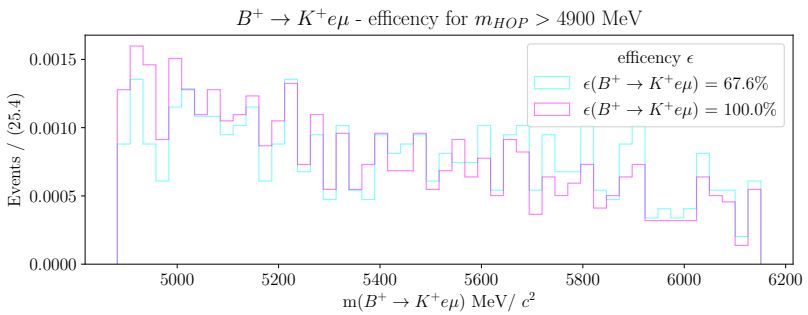

Figure 4.14: Normalised $B^+ \to K^+e^\pm\mu^\mp$ sample showing the impact of the HOP selection on a combinatorial shape - background shape remains unchanged due to this selection. Purple shows the full sample without any m_{HOP} selection, blue shows the sample passing $m_{HOP} > 4900\,\text{MeV}$. The shape is largely unchanged, and easily modelled with an exponential. Where efficiency refers to the accept/reject percentage for a $m_{HOP} > 4900\,\text{MeV}$ selection.

constant as a function of m_{Kee} or $m_{Ke\mu}$, both the m_{Kee} and $m_{Ke\mu}$ distributions can be described with a falling exponential and yield a good quality of fit. Additionally, the efficiencies of 72.0% for Run1 and 68.7% for Run2 are comparable with the efficiency of the $B^+ \to K^+e^\pm\mu^\mp$ sample, showing consistent performance across both runs and two types of data, giving confidence in the robustness of the selection.

The impact of the m_{HOP} cut on the control mode and partially reconstructed m_{Kee} distributions is shown in Fig. 4.16. For the control mode in the data, the m_{HOP} cut has an efficiency of 86% and 87% on the signal simulated samples for Run1 and Run2 respectively. The impact on the partially-reconstructed background is studied using $B^{+,0} \to J/\Psi(e^+e^-)K^{*+,0}$ simulation data where the m_{HOP} cut has an efficiency of 50% and 50% for Run1 and Run2 MC respectively.

Similarly, the impact of the m_{HOP} cut on simulated rare mode signal and partially reconstructed background events is shown in Fig. 4.17. The signal efficiency of the m_{HOP} in the rare mode is 91% and 92% for Run1 and Run2 MC respectively. The efficiency of the m_{HOP} cut on simulated $B^{+,0} \to K^{*+,0}e^+e^-$ events is 51% and 51% for Run1 and Run2 MC respectively.

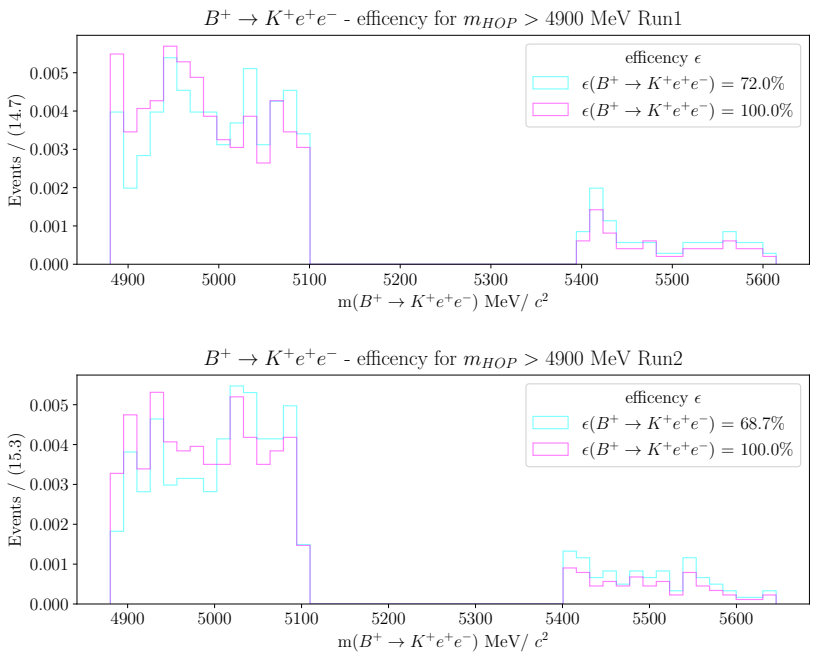

Figure 4.15: Normalised $B^+ \to K^+e^+e^-$ data sample for Run1 (top) and Run2 (bottom) showing the impact of the HOP selection on a combinatorial shape, with no selection on the combinatorial background to focus on HOP mass impact - the background shape remains unchanged due to this selection. Purple shows the full sample without any m_{HOP} selection, blue shows the sample passing $m_{HOP} > 4900\,\text{MeV}$. (Signal region blinded.) Where efficiency refers to the accept/reject percentage for a $m_{HOP} > 4900\,\text{MeV}$ selection.

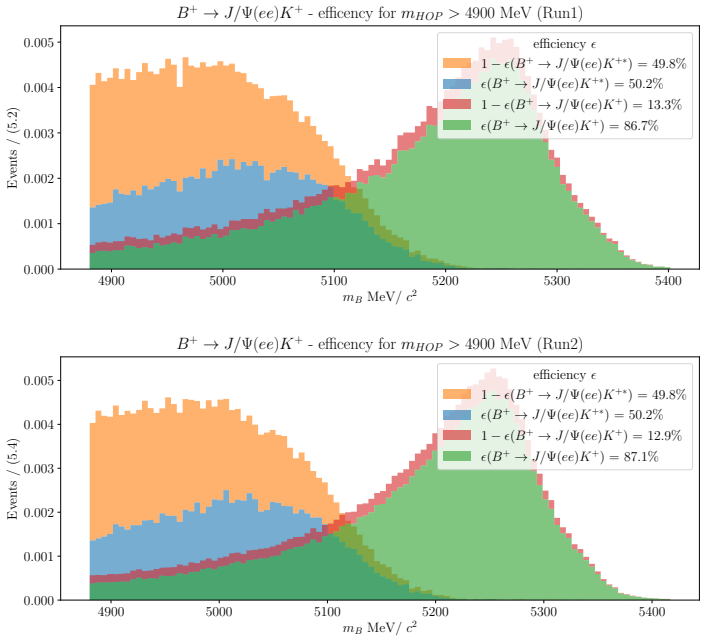

Figure 4.16: Stacked histograms showing the selection process on the $B^+ \to J/\Psi K^+$ and $B^+ \to J/\Psi K^*$ reconstructed mass distributions, Run 1 (top) and Run 2 (bottom). The stacked histograms show the distribution of events with respect to the m_{HOP}, the lower portion is $m_{HOP} > 4900$ MeV and the upper portion $m_{HOP} < 4900$ MeV. Where efficiency refers to the accept/reject percentage for a $m_{HOP} > 4900\,\text{MeV}$ selection.

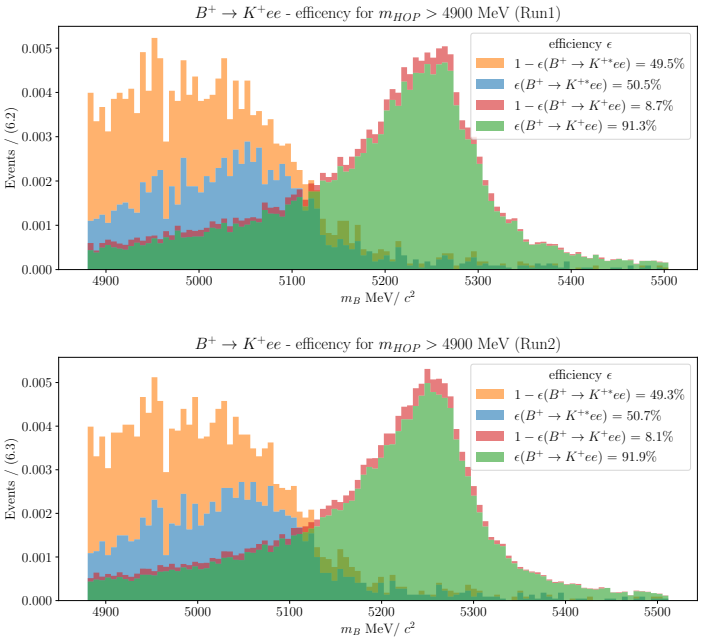

Figure 4.17: Stacked histograms showing the selection process on the $B^+ \to K^+e^+e^-$ and $B^+ \to K^*e^+e^-$ reconstructed mass distributions, Run 1 (top) and Run 2 (bottom). The stacked histograms show the distribution of events with respect to the m_{HOP}, the lower portion is $m_{HOP} > 4900$ MeV and the upper portion $m_{HOP} < 4900$ MeV. Where efficiency refers to the accept/reject percentage for a $m_{HOP} > 4900\,\text{MeV}$ selection.

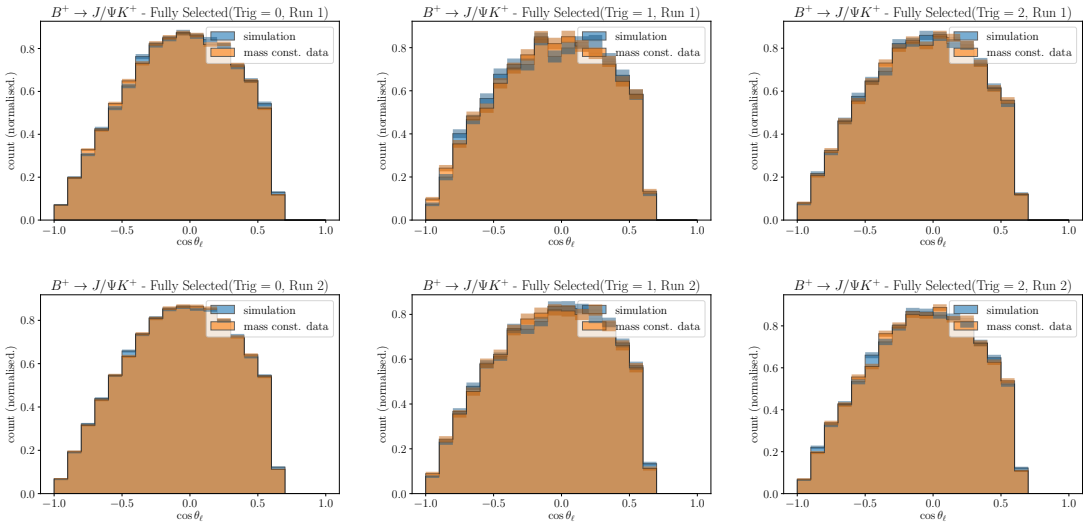

Figure 4.18: Control mode fully selected, with a selection on the constrained mass, simulation (blue) and data (orange) angular distribution comparison. The results show that the selection process is well modelled within statistical uncertainty.

4.6.4 Final Offline Selection

In Table 4.9 the final selections are shown for Run1 and Run2, with the cuts placed on specific trigger categories specified, and relevant sub-regions identified. Further results presented in this chapter described as *"fully selected"* have these selections applied unless otherwise stated. To assess the final modelling of the selections on the angular distribution, the fully selected control mode data and simulation with a cut on the constrained mass applied are compared. Figure 4.18 shows the agreement between simulation and data split by trigger category and run, and is found to be in good agreement.

Table 4.9: Table showing the final fully selected criteria for the rare mode and control mode. These are applied to the already pre-selected tuples. The notation $m_{\ell\to\pi}$ indicates the reconstructed mass distribution under the hypothesis of swapping the electron mass for the pion mass.

Run	var.	cut	region
Run1	BDT eTOS Run1	> 0.89	eTOS
	BDT Run1	> 0.79	hTOS
	BDT Run1	> 0.87	TIS
	m_{HOP}	> 4900 MeV	
	lepton-kaon θ	> 0.001	
	$m(K^+\ell^-)$	$> m(D^0)$	
	$\lvert m_{\ell\to\pi}(K^+\ell) - m(D^0)\rvert$	$> 40\,\text{MeV}$	
Run2	BDT eTOS Run2	> 0.86	eTOS
	BDT Run2	> 0.86	hTOS
	BDT Run2	> 0.85	TIS
	m_{HOP}	> 4900 MeV	
	lepton-kaon θ	> 0.001	
	$m(K^+\ell^-)$	$> m(D^0)$	
	$\lvert m_{\ell\to\pi}(K^+\ell) - m(D^0)\rvert$	$> 40\,\text{MeV}$	

4.7 Mass Fits

Extended un-binned maximum likelihood fits are performed to the reconstructed mass distribution m_{Kee} for both $B^+ \to J/\Psi(e^+e^-)K^+$ and $B^+ \to K^+e^+e^-$ candidates to obtain the yields of the signal and background components. Precise values for these yields allows the angular fit to be performed on background subtracted data. The backgrounds obtained by fitting m_{Kee} cannot be weighted away from the angular distribution using an ${}_s\mathcal{W}$eight method due to the correlation between m_{Kee} and $\cos\theta_\ell$[34]. Instead the background yields are used to independently subtract angular templates of the background shapes. The PDF that describes the reconstructed mass shape in each trigger category can generically be given as

$$F^{\text{trig}}(m) \propto N_{\text{sig}}^{\text{trig}} S^{\text{trig}}(m) + N_{\text{bkg}}^{\text{trig}} B^{\text{trig}}(m), \tag{4.7}$$

where S^{trig} and B^{trig} describe the signal and background PDFs for that trigger category, while $N_{\text{sig}}^{\text{trig}}$ and $N_{\text{bkg}}^{\text{trig}}$ are the yields obtained from the fit. The signal line shape depends on the bremsstrahlung recovery process, and specifically whether the energy from photon clusters has been recombined with: either, none, or both of the electrons. These bremsstrahlung categories are defined as follows:

- 0γ - no photons recovered, energy of neither electron is corrected;
- 1γ - one electron has its energy corrected;
- 2γ - both electrons have their energy corrected.

These three bremsstrahlung categories are combined to make the full signal PDF as

$$\mathcal{F} = f_{0\gamma}^{\text{trig}} S_{0\gamma}^{\text{trig}}(m) + f_{1\gamma}^{\text{trig}} S_{1\gamma}^{\text{trig}}(m) + (1 - f_{0\gamma}^{\text{trig}} - f_{1\gamma}^{\text{trig}}) S_{2\gamma}^{\text{trig}}(m) \tag{4.8}$$

where $f_{0\gamma}^{\text{trig}}$ and $f_{1\gamma}^{\text{trig}}$ are the fractions of 0γ and 1γ events and are constrained to values from simulation.

Due to the strong dependence of the signal shape on the photon recovery each bremsstrahlung category has its own empirical model obtained from simulated samples.

In the case of 0γ, where no photon cluster is added to the electron momentum, the line shape is characterised by a Gaussian core with a long power-law tail to low masses, due to the missing energy from bremsstrahlung. Modelling with a double sided crystal ball function [44] gave the best description of the distribution. In the cases of either 1γ or 2γ where some momentum is recovered the asymmetry is reduced, however, to model the increased width an additional Gaussian core is added to the double sided crystal ball due to incorrectly added photon clusters. For each bremsstrahlung category the mean is shared between the left and right side crystal balls, as well as the additional Gaussian in 1γ or 2γ, to ensure a mono-modal distribution. Floating the large number of parameters in the low statistics rare mode would result in an unstable fit, so the control mode and simulated samples are used to obtain the description of the signal shape which is then fixed when fitting the experimental data. Per run period there are nine different signal sub-models, one for each of the three bremsstrahlung recovery cases, in each of the three trigger categories. In each trigger category individual line shapes are obtained for each bremsstrahlung category from simulated samples, and fixed in the final mass fit.

4.7.1 Determining Signal Line Shapes

To obtain the individual bremsstrahlung line shapes of these nine separate trigger and bremsstrahlung combinations in the control mode a simultaneous fit is performed using simulated samples and control mode full selected data with a selection on the constrained mass to obtain the line shapes. For the rare mode, only the simulated samples are fitted, and a correction factor is applied taken from the control mode fit.

The PDFs fitted to data and simulation share all parameters, however additional scaling terms for the mean $\Delta\mu^{\text{trig}}_{N\gamma}$, and the widths $\Delta\sigma^{\text{trig}}_{N\gamma}$ accommodate the variation between data and simulation. This scaled PDF is used to fit the final control mode data, and the scaling terms obtained in the control mode are used in the rare mode fit. The ratio of bremsstrahlung recovery categories in the final fit is taken from simulation.

The resulting fits to the control mode that determine the signal line shape parameters

can be found in Figs. [A.7 - A.12], and show that the data is well modelled by the simulation and appropriate small transformation. From these results it can be surmised that the singular fit to the rare mode simulation samples will provide an adequate model of the rare mode data once the appropriate corrections to the m_{Kee} line-shape have been applied, which are taken from the control mode. The corrections can be translated from the control mode to the rare mode, as it was found that the core of the m_{Kee} line-shape is largely independent of q^2.

4.7.2 Fit Model to Data

When fitting the data, each of the background components needs an individual PDF to describe it. The combinatorial background contribution is modelled with a falling exponential, with decay constant τ_{trig}, for each trigger category, data-taking period, and m_{HOP} region. The partially reconstructed background is dominated by $B^{+,0} \to K^{*+,0}e^+e^-$ and $B^{+,0} \to J/\Psi(e^+e^-)K^{*+,0}$ decays in the rare mode and control mode respectively, each is modelled using a kernel density estimation (KDE) from simulation samples. In the fitting procedure each KDE is obtained from a fit to simulation samples for that trigger category, data taking period, and m_{HOP} region.

The dataset is split into two regions, a low m_{HOP} region ($m_{HOP} < 4900\,\text{MeV}$) and a high m_{HOP} region ($m_{HOP} > 4900\,\text{MeV}$) for each of the three exclusive trigger categories. This selection splits the data into a partially reconstructed enriched region (low m_{HOP}) and depleted (high m_{HOP}) regions with respect to the signal yield. In this scheme a six part simultaneous fit is performed in each run to the three exclusive trigger categories split into two m_{HOP} regions such that a single yield for the partially reconstructed events can be floated. The relative yield in each of the fits is scaled appropriately using simulation samples of $B^{+,0} \to J/\Psi(e^+e^-)K^{*+,0}$ decays for the control mode and $B^{+,0} \to K^{*+,0}e^+e^-$ decays for the rare mode. This approach exploits the full statistical power of the dataset to float a single variable in a physically meaningful way that ensures a realistic distribution of partially reconstructed events, and provides an additional constraint

to the signal and combinatorial yields as well by reducing the freedom of the partially reconstructed background. In the control mode, the data sample is large enough that the signal yield for each trigger category is floated independently, while still being split into two m_{HOP} regions. The low region has three components to the fit, the partially-reconstructed background (modelled using a KDE fitted to $B^{+,0} \to J/\Psi(e^+e^-)K^{*+,0}$ and $B^{+,0} \to K^{*+,0}e^+e^-$ simulation samples for the control mode and rare mode respectively), the signal leaking into the region (modelled using a KDE fitted to $B^+ \to J/\Psi(e^+e^-)K^+$ and $B^+ \to K^+e^+e^-$ simulation samples for the control mode and rare mode respectively), and remaining combinatorial background (modelled as a falling exponential).

4.7.3 Control Mode Fit Results

The mass fits are initially performed in the control mode to validate the fitting procedure for the rare mode. It should be noted that the model used is not meant to be a perfect description of the control mode data, but an adequate one considering that the rare mode will have 100 times fewer events. The mass fits to simulated $B^+ \to J/\Psi(e^+e^-)K^+$ samples are used to fix the shape of the signal PDF in the fit to data, these fits are performed in individual bremsstrahlung and trigger categories and can be found in the appendix in Figs. [A.7 - A.12]. The fit results to the low m_{HOP} region can be found in Fig. 4.19 , showing the signal component leaking into the low region, the residual combinatorial background, and the partially-reconstructed component. The signal and partially reconstructed shapes are taken from simulated samples using a KDE. The fits to the high m_{HOP} region can be found in Fig. 4.20 , showing the signal component and the three bremsstrahlung categories, the small combinatorial component, and the partially-reconstructed component.

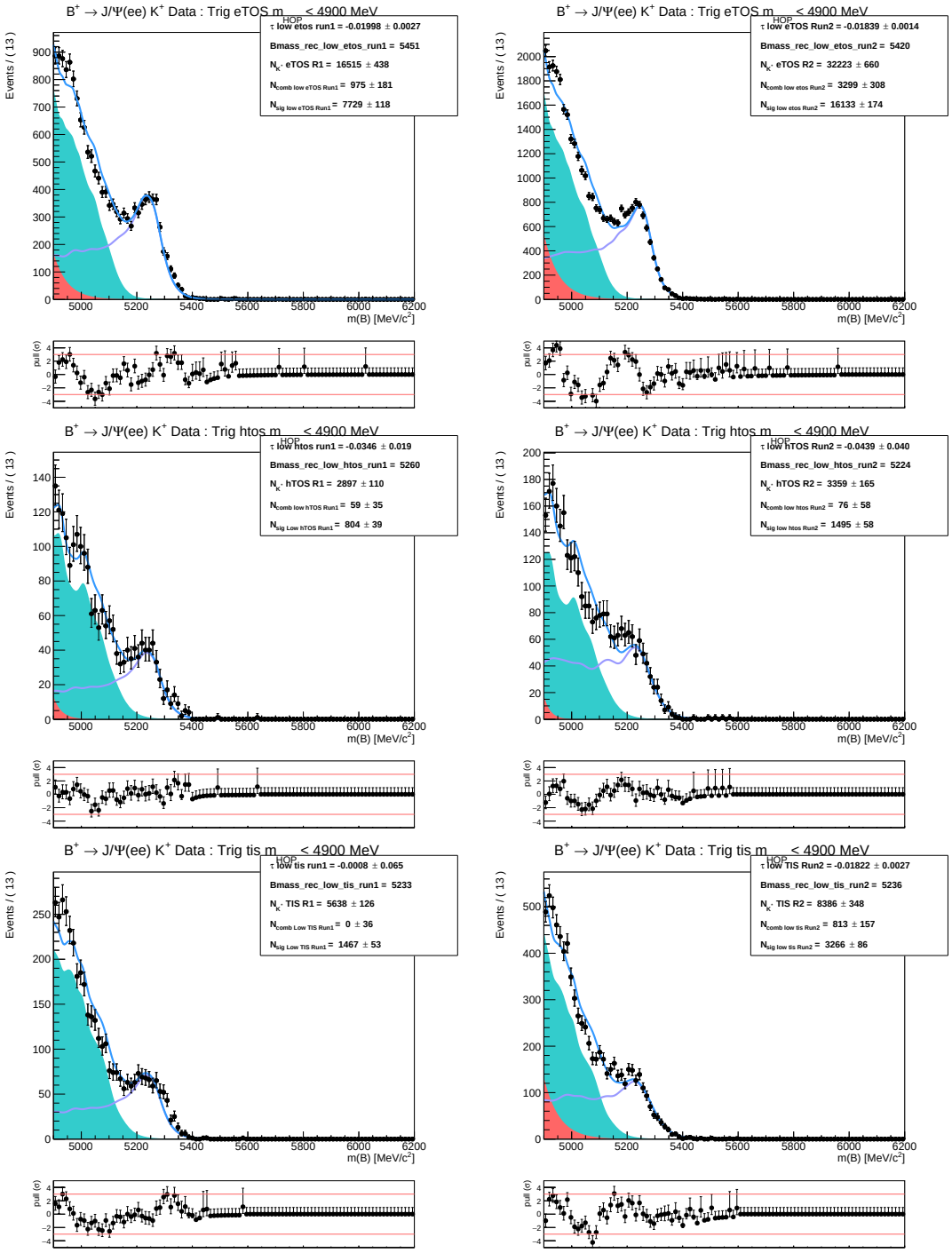

Figure 4.19: Run1 (left) and Run2 (right) fits to the $m_{HOP} < 4900\,\text{MeV}$ region for $B^+ \to J/\Psi(e^+e^-)K^+$ fully selected data, split into trigger categories $eTOS$ (top), $hTOS$ (centre), TIS (bottom). The combinatorial background is shown in red, the signal leaking into the low HOP region in purple, the partially reconstructed background in teal, and the total model in blue.

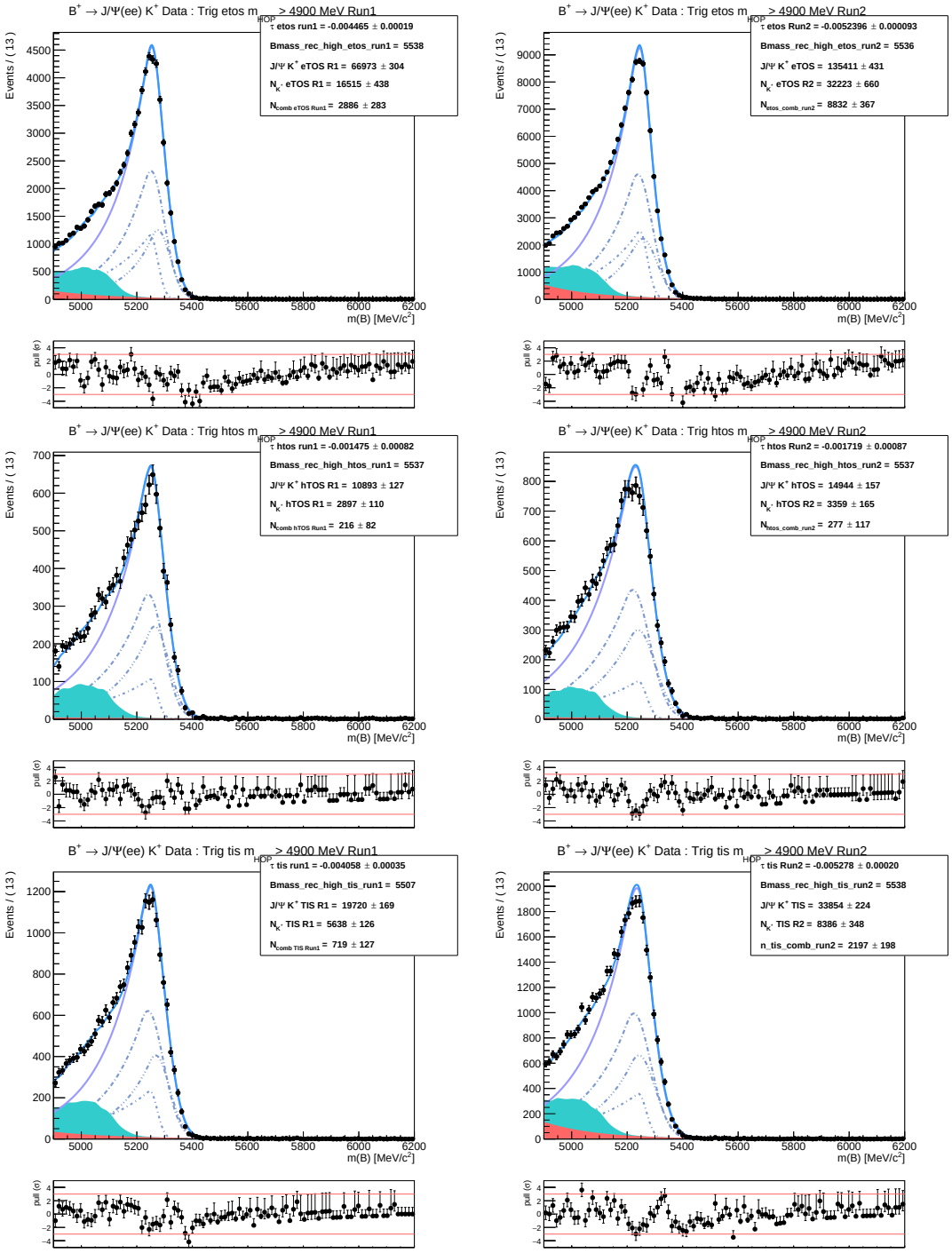

Figure 4.20: Run1 (left) and Run2 (right) fits to the $m_{HOP} > 4900\,\text{MeV}$ region for $B^+ \to J/\Psi(e^+e^-)K^+$ fully selected data, split into trigger categories $eTOS$ (top), $hTOS$ (centre), TIS (bottom). The combinatorial background is shown in red, the partially reconstructed background in teal, and the entire signal model in purple. Each individual bremsstrahlung category as a dashed blue line, and the total model as a solid blue line.

Trigger	$N_{B^+\to K^+e^+e^-}$	$N_{\text{part-reco}}$	N_{comb}
$eTOS$	201.7 ±17.3	57.5 ±12.9	23 ±16
$hTOS$	69.5 ±6.0	24.5 ±12.9	99 ±15
TIS	73.0 ±6.3	24.8 ±5.6	38 ±12
Total	348 ±30	185 ±29	150 ±43

Table 4.10: Mass fit component yields for $B^+ \to K^+e^+e^-$ decays in Run1. The signal and part-reco yields in individual trigger categories are derived quantities from the total yields.

4.7.4 Rare Mode Fit Results

The rare mode fit broadly follows the same method outlined in Sec. 4.7. The main difference is that each individual bremsstrahlung category is fitted in the fully selected simulation data, but unlike the control mode the fit is not done simultaneously with data. The rare mode simulation fits in each bremsstrahlung category can be found in Figs.[A.15 - A.20]. Taking each of these bremsstrahlung categories and combining them for each trigger the simulation is fitted again to show that the combination is correctly modelled, fits to each trigger category can be found in Fig. A.21 and Fig. A.22. The corrections to the mean and width of the mass fit distributions obtained from the control mode in Sec. 4.7 are applied to the rare mode fit. The fits to the rare mode use the same m_{HOP} scheme to simultaneously fit the high and low regions across all three trigger categories to control the part-reco background, and the results for the low m_{HOP} region can be seen in Fig. 4.21, with components showing the leaking signal from the high region, part-reco component, and the minimal residual combinatorial background. The results of the fit to the rare mode exclusive trigger categories can be seen in Fig. 4.22. These plots show the components for the signal fit as a combination of bremsstrahlung categories, the constrained part-reco component, and the combinatorial background. The yields for each component of the fit to the rare mode data are shown in Tab. 4.10 and Tab. 4.11 for Run1 and Run2 respectively.

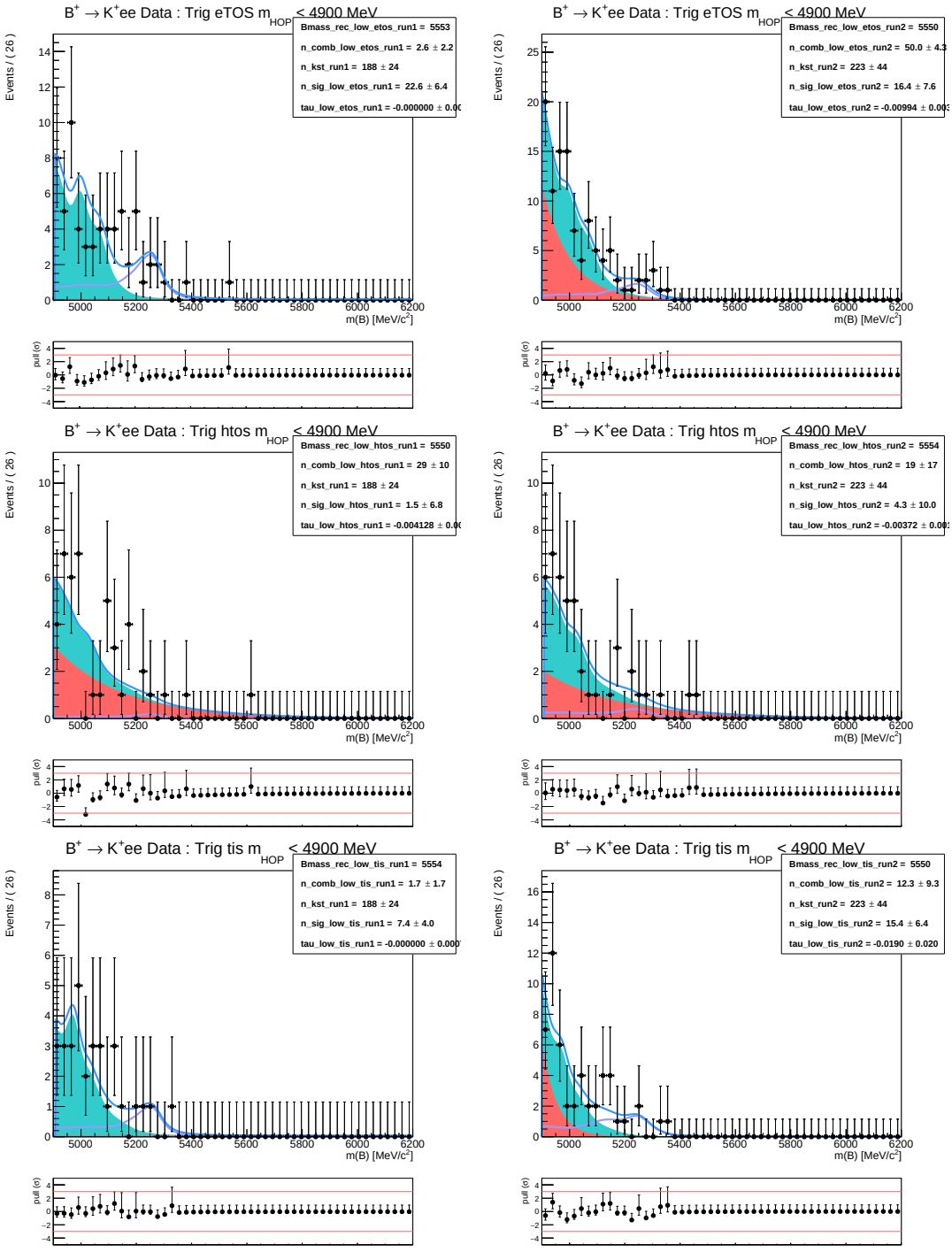

Figure 4.21: Run1 (left) and Run2 (right) mass fits to the $m_{HOP} < 4900\,\text{MeV}$ region for $B^+ \to K^{++}e^-$ fully selected data, split into trigger categories $eTOS$ (top), $hTOS$ (center), TIS (bottom). The combinatorial background is shown in red, the signal leaking into the low HOP region in purple, the partially reconstructed background in teal, and the total model in blue.

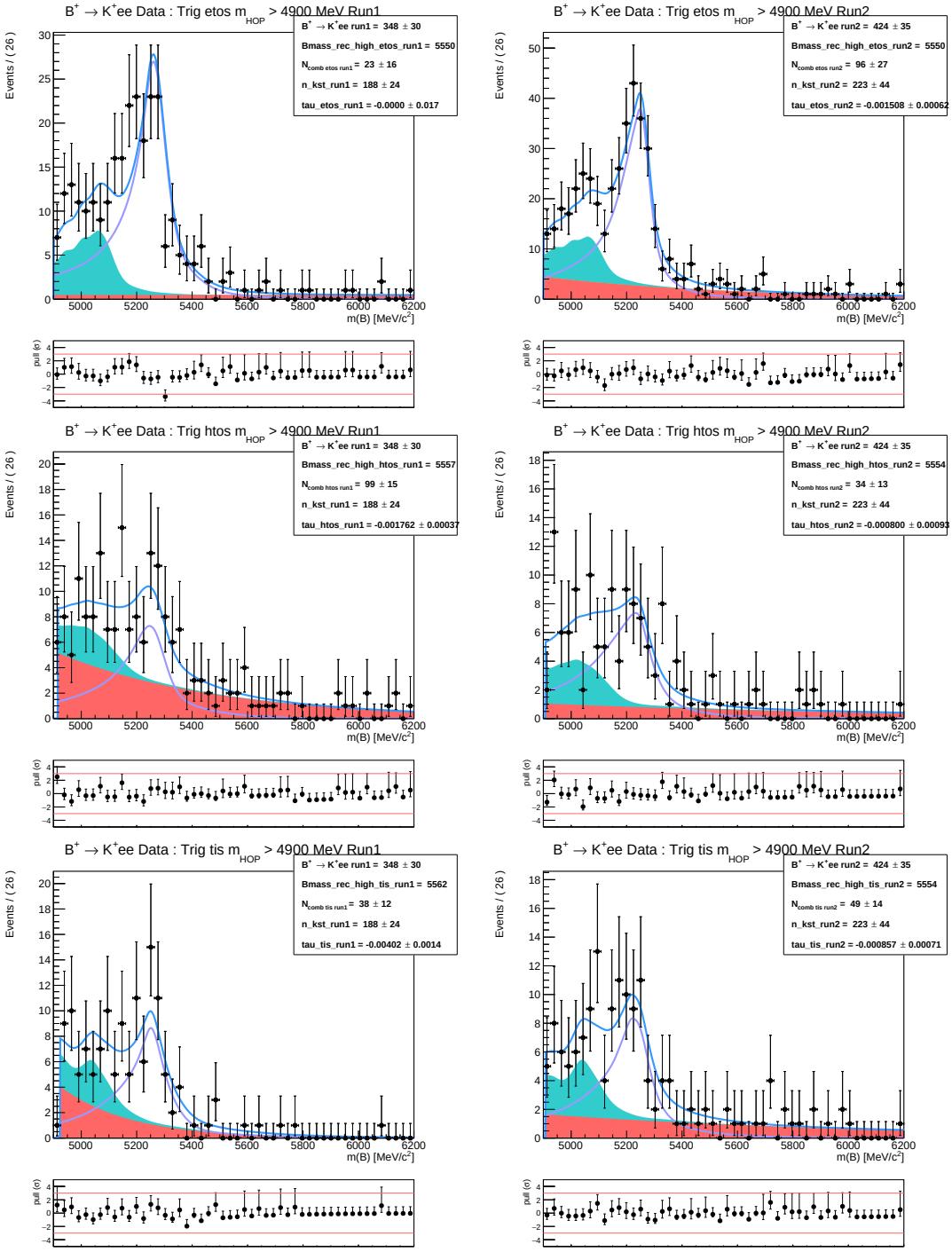

Figure 4.22: Run1 (left) and Run2 (right) mass fits to the $m_{HOP} > 4900\,\text{MeV}$ region for $B^+ \to K^+e^+e^-$ fully selected data, split into trigger categories $eTOS$ (top), $hTOS$ (centre), TIS (bottom). The combinatorial background is shown in red, the partially reconstructed background in teal, and the entire signal model in purple. Each individual bremsstrahlung category as a dashed blue line, and the total model as a solid blue line.

Trigger	$N_{+B\to K^+e^+e^-}$	$N_{\text{part-reco}}$	N_{comb}
$eTOS$	266.6 ±21.7	68.3 ±24.2	96 ±27
$hTOS$	80.3 ±6.5	29.0 ±24.2	34 ±13
TIS	76.5 ±6.2	29.5 ±10.5	49 ±14
Total	424 ±35	226 ±45	173 ±55

Table 4.11: Mass fit component yields for $B^+ \to K^+e^+e^-$ decays in Run2. The signal and part-reco yields in individual trigger categories are derived quantities from the total yields.

4.7.5 Background Subtraction

In this analysis the measurement of the angular observables is made as an average in a true q^2 bin. This means that the data must be unfolded to account for migrations between q^2 bins, however the signal events and background events will not necessarily migrate in the same way. The background contributions for the partially reconstructed could be modelled and unfolded using simulated samples, however the combinatorial can not. Therefore, prior to unfolding the backgrounds are subtracted using two dimensional templates in reconstructed q^2 and $\cos\theta_\ell$, and the result is unfolded. The combinatorial background template is modelled using real data from the upper mass side band with a relaxed BDT cut to ensure a well populated distribution free from noise artefacts. This had no effect on the shape of the background angular template. The partially reconstructed template comes from the fully selected $B \to K^*e^+e^-$ simulation data, where the statistics are high enough to provide a smooth distribution with all selections applied. The background templates (combinatorial and part-reco) used in the subtraction are scaled to the yields obtained for each background component from the mass fit, and the templates are taken as 2D histograms in 20 bins of $\cos\theta_l$ and 3 bins of q^2. The templates for the combinatorial and the part-reco backgrounds can be found in Fig. 4.23. The independence of the background $\cos\theta_l$ distribution on BDT selection was validated. The same distributions for the central q^2 bin, $2 < q^2 < 5$ GeV, is shown in Fig. 4.24.

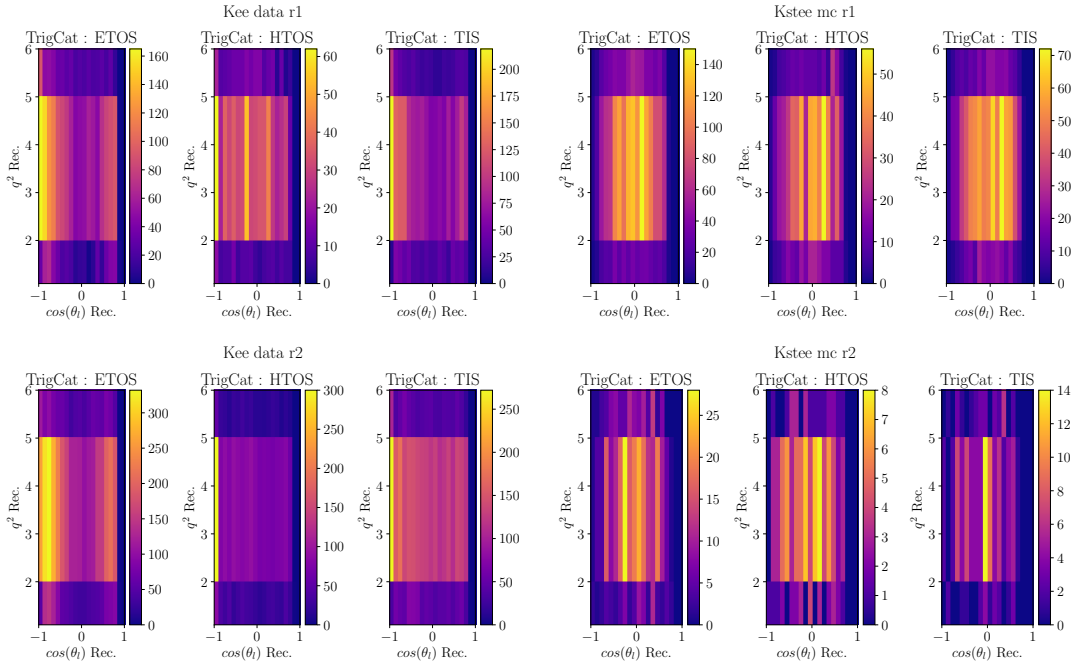

Figure 4.23: Run1 (top) and Run2 (bottom) background templates in three bins of $q2$ and twenty bins of $\cos\theta_l$ from the upper mass side band data (left), and the $B \to K^* e^+ e^-$ simulation samples (right). Selection made for the upper mass side band with a relaxed cut on the BDT to smoothly model the shape of the background across $\cos\theta_l$.

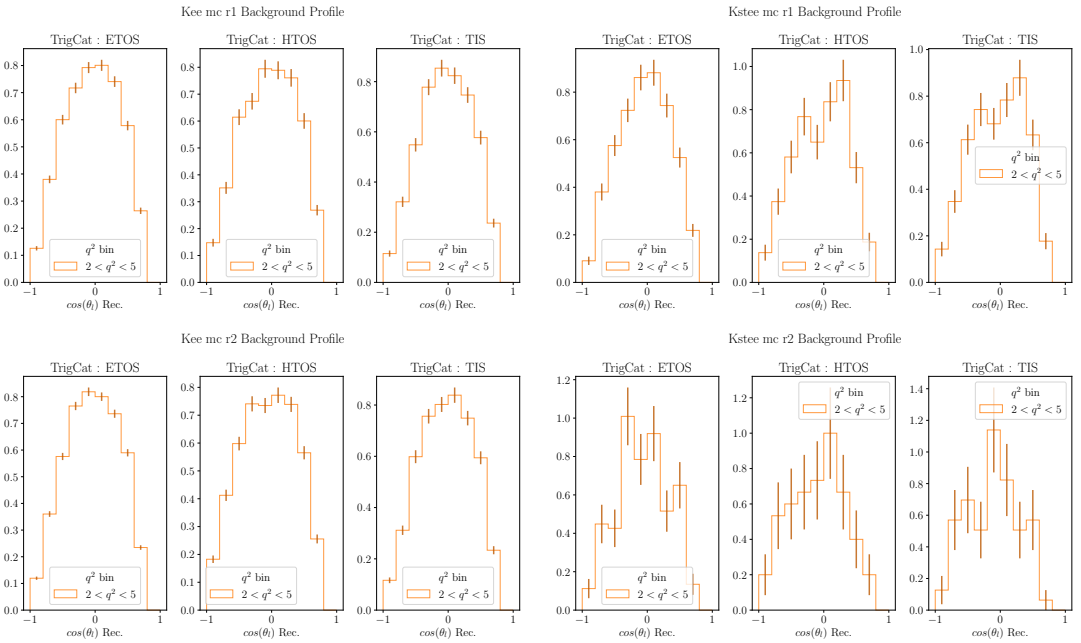

Figure 4.24: Run1 (top) and Run2 (bottom) background templates in three bins of $q2$ and twenty bins of $\cos\theta_l$ from the upper mass side band data (left), and the $B \to K^* e^+ e^-$ simulation samples (right). Selection made for the upper mass side band with a relaxed cut on the BDT to smoothly model the shape of the background across $\cos\theta_l$.

4.8 Unfolding

The aim of the analysis is to measure the observables A_{FB} and F_H as q^2 averages $\langle A_{FB}(q^2)\rangle$ and $\langle F_H(q^2)\rangle$ in the true q^2 bin of $q^2 \in [2, 5]$ GeV. When measuring a parameter of a decay a reconstructed value is obtained, which by definition is only an approximation of the true value. In the case of a binned analysis it is important to understand what proportion of events are reconstructed in a bin that contains the true value for that event. The resolution effects that cause events to potentially be reconstructed in the wrong bins are called migrations. By comparing q^2_{true} and q^2_{rec} in three bins of q^2 [1.1 - 2.0 - 5.0 - 6.0] the migrations between q^2 bins can be seen. In Fig. 4.25 this effect is shown for Run1 and Run2 simulation rare mode samples, and the overall picture can be understood. The majority of events remain on the diagonal of the matrix and are therefore reconstructed in the appropriate q^2 bin (89.2% in both Run1 and Run2.) No migrations are seen that span more than a whole bin in the q^2 binning scheme, which shows that the 1 GeV buffer acts as expected and migrations from outside the q^2 range 1.1 - 6 GeV can be safely ignored.

As the $\cos\theta_l$ distribution is dependent on q^2, the q^2 distribution cannot be unfolded independently of $\cos\theta_l$. As the migrations in $\cos\theta_l$ are small, and given the limited simulation data, the angular distribution was not unfolded. However, a quasi diagonal unfolding matrix for q^2 was used to preserve the $\cos\theta_l$ dependence on q^2 once the unfolding was performed. This allows four variables to be encoded in a two-dimensional matrix, the structure can be seen in Fig. 4.26. The two dimensional matrix of $\cos\theta_l$ in three bins of q^2 [1.1 - 2.0 - 5.0 - 6.0], becomes a vector [$\cos\theta_l(1.1 < q^2 < 2.0)$, $\cos\theta_l(2.0 < q^2 < 5.0)$, $\cos\theta_l(5.0 < q^2 < 6.0)$], where no transitions between bins of $\cos\theta_l$ are considered. This results in a response matrix with a tri-diagonal structure with no transitions between $\cos\theta_l$ bins. As such, the transition probability kernel describes the probability that a true event in a specific $\cos\theta_l$ bin is found in the same angular bin but a different kinematic bin. The sparse, diagonalised, 60×60 matrix seen in Fig. 4.26 describes the transition probabilities. This can be constructed as a familiar matrix inversion problem.

A Bayesian unfolding procedure is used to estimate a systematic uncertainty on the

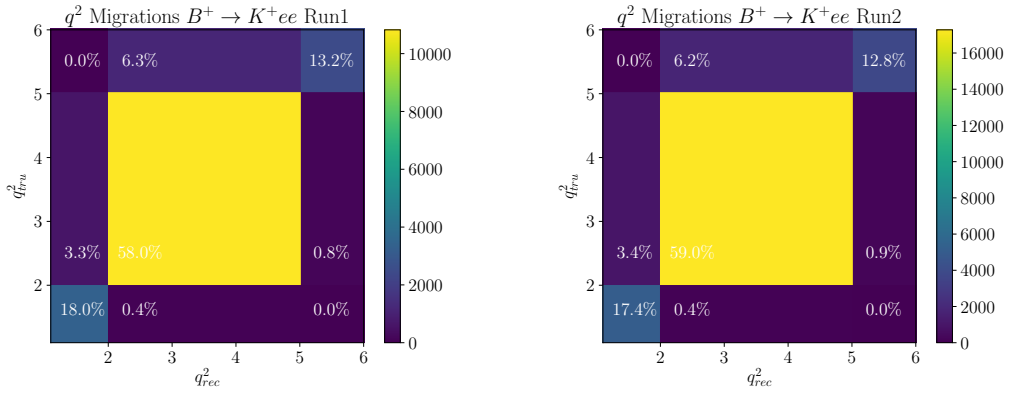

Figure 4.25: Matrix showing the percentage distribution of migrations between q^2_{true} and q^2_{rec}. Run1 (left) and Run2 (right) have very similar distributions. These matrices are shown normalised to 100% of all transitions as an overview of the entire set of migrations to emphasis the dominant regions. In Run1 58% (Run2 59%) of the events remain in the central q^2 bin, where as in Run1 6.3% (Run2 6.2%) of $2 < q^2_{true} < 5$ events are reconstructed in $5 < q^2_{rec} < 6$.

choice of the unfolding method. The binning scheme of 20 bins in $\cos\theta_l$ is motivated by the resolution study placing a lower limit on the $\cos\theta_l$ bin size of 0.1. The study of the resolution effects can be found in Sec. 4.5.4. The Python package `PynFold` is used to perform the unfolding [45].

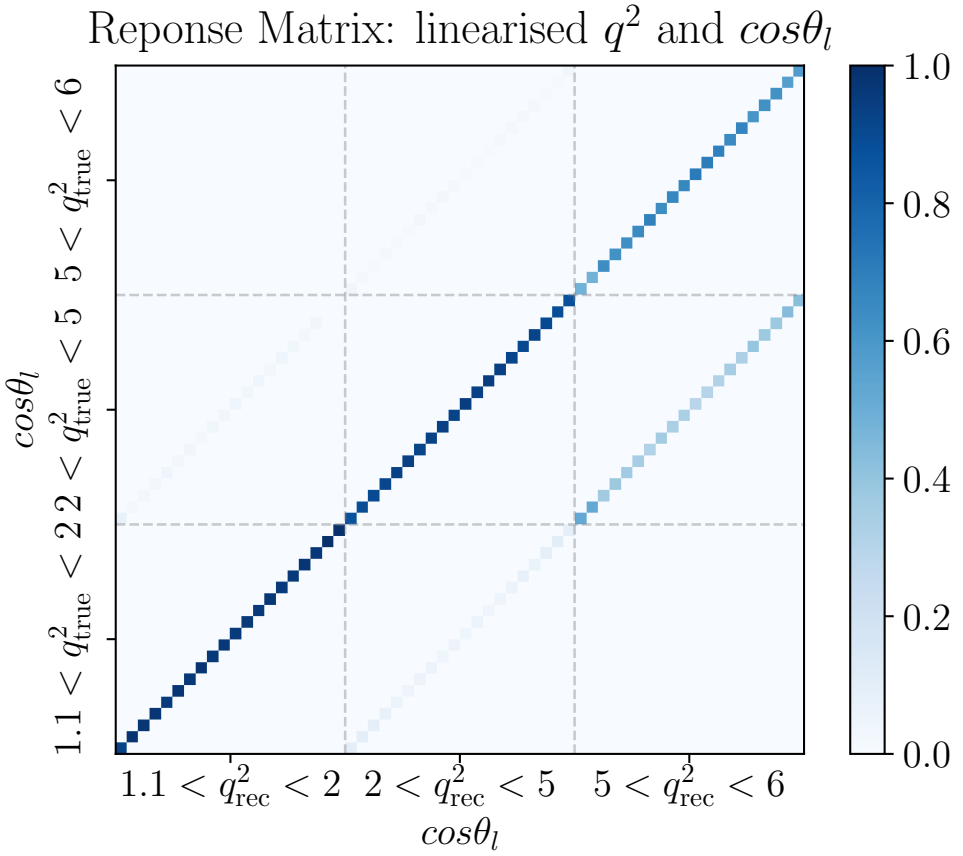

Figure 4.26: Unfolding matrix - showing the transition probability between reconstructed and true q^2 in bins of linearised $\cos_l$ in bins of q^2. The x-axis shows the binned true $\cos\theta_l$ distribution in three bins of reconstructed q^2, while the y-axis shows the binned true $\cos\theta_l$ distribution in three bins of true q^2. The probabilities that a measured $\cos\theta_l$ value in a q^2 bin is reconstructed in the three reconstructed q^2 bins can be read down each column. *Note: Entries outside the three diagonals are 0 by construction.*

4.9 Acceptance Corrections

The efficiency of reconstructing, and selecting $B^+ \to K^+e^+e^-$ decays, referred to as *acceptance*, is not uniform across the angular distribution and therefore has an impact on the final angular fit to extract the angular observables. To account for this lack of uniformity, the efficiency over $\cos\theta_l$ is modelled from the simulation and used to modulate the angular decay rate in the final fit. The efficiency losses can come from the geometric acceptance of the detector, triggers, stripping, and the selections applied. The angular acceptance correction is calculated on the fully selected simulated samples. The methodology to calculate the acceptance corrections is the same for both $B^+ \to J/\Psi(\to e^+e^-)K^+$ samples and $B^+ \to K^+e^+e^-$ samples.

The acceptance can be thought of as a fraction that describes what proportion of the angular distribution is measured at each point in $\cos\theta_l$, such that if the reconstruction was perfectly efficient it would be a flat line at 1. To model the full efficiency the generated angular distribution is compared to the fully reconstructed and selected simulation samples, in a true q^2 region.

More formally, the $\cos\theta_l$ distribution, in a $2 < q^2_{true} < 5\,\text{GeV}^2$ region, of simulated signal decays passing all selections with all correction weights applied is divided by the generator level distribution in the same q^2_{true} region. The generator level distribution simply means the base physics parabola that is used to generate events in the Monte Carlo. This is obtained by running `EvtGen`[46] standalone for this mode without any geometric or kinematic cuts, other than the q^2 selection.

The normalised efficiency is then calculated for each weighted simulation event as:

$$\epsilon_i(cos\theta_l) = \frac{w_i(cos\theta_l)}{d\Gamma/d\cos\theta_l}, \tag{4.9}$$

where $\epsilon_i(cos\theta_l)$ is the weighted efficiency per event i, $w_i(cos\theta_l)$ is the weight of each event i. The efficiency distribution in Eq. 4.9 describes the shape of the angular acceptance as a function of $\cos\theta_l$ at an arbitrary scale. Once the acceptance correction has been obtained it is used to modulate the decay rate given in Eq. 2.84. The real data can be fit using the two floating parameters of A_{FB} and F_H.

To model the shape of the acceptance correction, the efficiency distribution is fit using a combination of Legendre moments and polynomial fits. This construction allows an analytic integral in the fitting procedure, and gives more control over the freedom of the fit than a density estimate method. The acceptance correction is computed in two stages: first, without any selections applied to veto the cascading semi-leptonic backgrounds, where symmetry in the distribution discussed in Sec. 4.6.2 can be exploited to use only even order Legendre moments. Secondly, the selections against the cascading backgrounds are applied, and a second order polynomial is used to fit the region $0.62 < \cos\theta_\ell < 0.75$, resulting in a smooth step. The normalisation ensures these regions transition reasonably smoothly, however there is a sharp transition as the PDF changes description. Above $\cos\theta_\ell > 0.75$ the acceptance is 0. The final acceptance correction is therefore a three part continuous PDF.

Each trigger category was fit independently for each run, and through trial and error it was found that an even-only 8th order polynomial fit the $eTOS$ category well, while a 6th order even-only polynomial was fit to $hTOS$ and TIS. While these specific orders provide a better model of the acceptance correction, when evaluated as a systematic uncertainty it was found that changing the order used to generate the acceptance correction had a negligible impact on the angular fit. The full acceptance becomes the step-wise function, given in Eq. 4.12, that can be seen fitting the efficiency distribution in Figs. [4.27 - 4.28] for both the control mode and the rare mode in both Run1 and Run2 simulated samples, split by trigger categories.

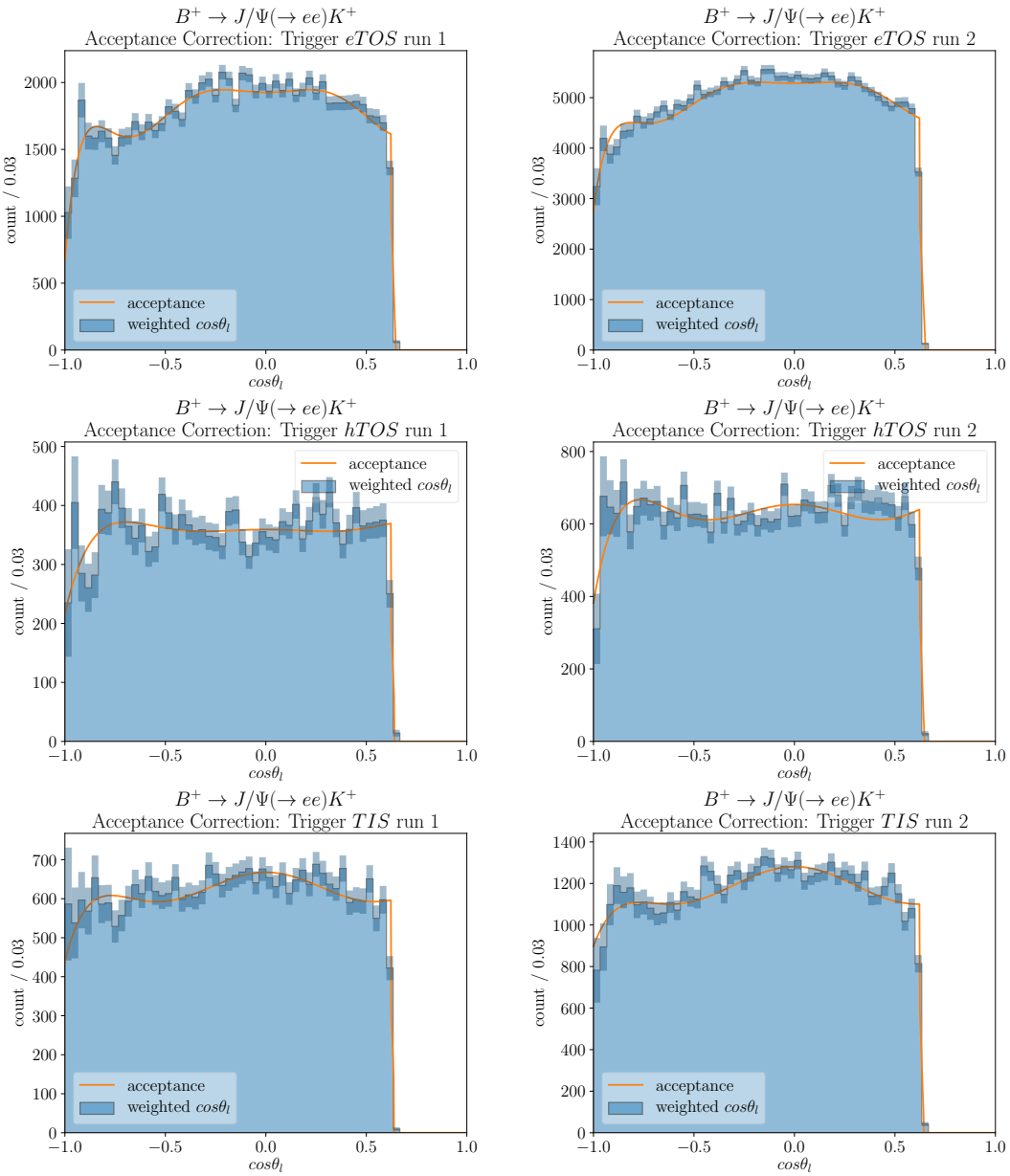

Figure 4.27: Acceptance corrections to $B^+ \to J/\Psi(e^+e^-)K^+$ simulation data for three exclusive trigger categories, eTOS (top), hTOS (centre), TIS (bottom), with semi-leptonic veto (Run1 - left) (Run2 - right). The binning granularity is much finer than in the angular fit. The acceptance correction is an 8th order (even only) Legendre polynomial for eTOS and a 6th order even only Legendre polynomial for hTOS and TIS.

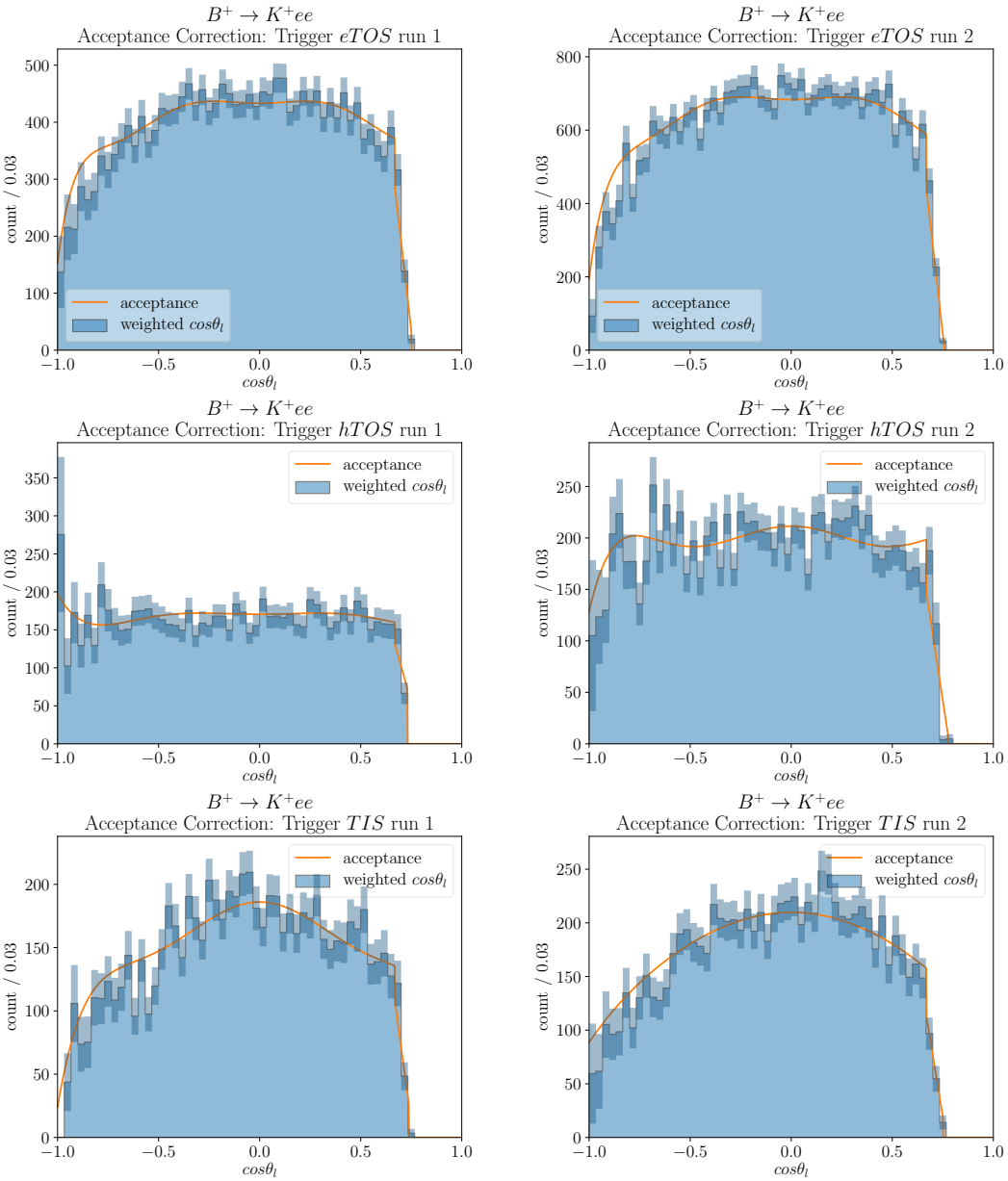

Figure 4.28: Acceptance corrections to $B^+ \to K^+e^+e^-$ simulation data for three exclusive trigger categories, eTOS (top), hTOS (centre), TIS (bottom), with semi-leptonic veto (Run1 - left) (Run2 - right). The binning granularity is much finer than in the angular fit. The acceptance correction is an 8th order (even only) Legendre polynomial for eTOS and a 6th order even only Legendre polynomial for hTOS and TIS.

4.10 The angular fit

The fit to obtain the angular observables A_{FB} and F_H is a binned maximum likelihood fit with 20 bins of $\cos\theta_l$ in a true q^2 bin of $2 < q^2_{true} < 5\,\mathrm{GeV}^2/c^4$. The choice of the number of bins was motivated by the resolution in $\cos\theta_\ell$. The final result is given as a likelihood surface. The following section is split into several key parts: the angular model itself; the minimisation procedure; the Feldman Cousins technique to evaluate the statistical uncertainties in the fit; the fit to the control mode; the fit using toys of the rare mode; and finally the angular fit to the rare mode data.

4.10.1 Angular Model

The angular distribution of the $B^+ \to K^+e^+e^-$ decays is described in Eq. 2.84. Experimentally, this expression must be modified by the acceptance correction to account for the selection criteria, resulting in a full angular model given as

$$\int_{q^2_{min}}^{q^2_{max}} \frac{d^2\Gamma}{dcos\theta_l dq^2} \times \left\langle \varepsilon^{trig}_{eff}(cos\theta_l) \right\rangle_{q^2 bin} dq^2. \tag{4.10}$$

Where $\left\langle \varepsilon^{trig}_{eff}(cos\theta_l) \right\rangle_{q^2 bin}$ is the q^2 averaged acceptance correction for each exclusive trigger category, and is given by:

$$\left\langle \varepsilon^{trig}_{eff}(cos\theta_l) \right\rangle_{q^2 bin} = \begin{cases} \sum_{i=0,2,4,\ldots}^{O=6,8} L_i P_i(\cos\theta_l) & \text{for } \cos\theta_\ell < 0.62, \\ a\cdot(\cos\theta_l)^2 + b\cdot(\cos\theta_l) + c, & \text{for } 0.62 \le \cos\theta_\ell < 0.75, \\ 0. & \text{for } \cos\theta_\ell \ge 0.75 \end{cases} \tag{4.11}$$

where L_i are the Legendre moments, $P_i(\cos\theta_l)$ the Legendre polynomial and (a, b, c) the coefficients of the second order polynomial.

This fit results in measurements of $< A_{FB} >$ and $< F_H >$, which denote averages of A_{FB} and F_H in the region $2 < q^2_{true} < 5\ \mathrm{GeV}^2/c^4$. For the angular expression $\frac{1}{\Gamma_l}\frac{d\Gamma_l}{dcos\theta_l}$ to remain physical it must be positive definite for all values of $\cos\theta_l$ and as a result requires boundary conditions on the angular observables, which are given in Eq. 2.85. This

requirement has implications on the minimisation algorithm discussed in the following section.

4.10.2 Minimisation

This is a binned negative log-likelihood fit over 20 bins in $\cos\theta_\ell$ spanning -1 to 1. The expression in Eq. 4.10, which is the physics angular distribution modulated by the acceptance correction function, is integrated in each bin. The background subtraction method results in a weighted histogram where the sum of weights and sum of weights squared in each bin are not the same. There is also the potential for a negative sum of weights, which is not well described in the standard negative log-likelihood function. Therefore, the standard negative log-likelihood function using in a binned likelihood fit needed to be adapted.

The likelihood function used in the minimisation follows from Bohm's parameter estimation using data distorted by resolution effects [47]. For i events in bin j the sum of weights and the sum of weights squared are given as $n_j = \sum_i w_i$ and $w_j^2 = \sum_i w_i^2$. A scale factor is calculated per bin from the ratio of the two as $s_j = n_j/w_j^2$. In the data $n_j = w_j^2$, however this is not the case for the background subtraction templates as the templates are both scaled and negative. The fitting function is integrated and scaled in each bin by the total sum of weights as $\mu'_j = s_j \cdot n_j \int_{\text{bin}} f(x)dx$, and the sum of weights in each bin scaled as $n'_j = n_j \cdot s_j$. The sum of all scaled weights from all bins is given by $N = \sum_j n'_j$. The fitting function integrated over each bin where $n_j = 0$ is given as $\mu_j = \int_{\text{bin}} f(x)dx$. The likelihood function in Eq. 4.12 is given for three exclusive cases; positive sum of weights per bin, negative sum of weights per bin, and empty bins. For positive sum of weights, the likelihood is given by the difference between the function over the bin and the weights in the bin, with an additional negative log penalty, summed over all bins meeting the criteria. For negative sum of weights a Gaussian penalty term is given, summed over all bins meeting the criteria. For empty bins the expectation from the integral of the function contributes a penalty to the likelihood. The full likelihood

function over all j bins is given by:

$$\mathcal{L} = \begin{cases} -N + \sum_j \mu'_j - n'_j \cdot \log\left(\frac{\mu'_j}{n'_j}\right), & \text{for } n_j > 0 \text{ and } w_j^2 > 0, \\ \sum_j \frac{1}{2}\frac{(n_j-\mu_j)^2}{w_j^2}, & \text{for } n_j < 0 \text{ and } w_j^2 > 0, \\ \mu_j, & \text{for } w_j^2 = 0. \end{cases} \tag{4.12}$$

Due to the physical boundary conditions of the angular distribution discussed in Sec. 4.10.1, the minimisation of the angular fit cannot be done in a trivial way with naive gradient descent methods. Given the simplicity of the angular distribution of the signal and the fact that the background has been subtracted beforehand, the approach taken to deal with the boundary conditions is to step through F_H in small increments, float A_{FB} within the allowed range, and find the negative log-likelihood minimum of the ensemble of slices in F_H.

To leverage the maximum information from the data and to reduce the statistical uncertainties, the fit is performed as a single fit that combines the three triggers categories $eTOS$, $hTOS$, and TIS, and both Run1 and Run2 together. This is useful in the rare mode to minimise the number of potential negative bins in the angular fit in both the data and the Feldman Cousins evaluation. While Eq. 4.12 provides a protection against negative bins, combining trigger categories minimises the chance of observing negative weights. The same procedure is used for both the control mode and the rare mode to validate the method, even though the control mode is not statistically limited.

4.10.3 Statistical Uncertainty

The leading uncertainty on the angular observables comes from the limited amount of data in the rare mode. To estimate the statistical uncertainty, confidence intervals are estimated using a two-dimensional Feldman Cousins method [48].

This consists of uniformly sampling over the allowed (A_{FB}, F_H) space and calculating a test statistic at each location. For each (A_{FB}, F_H) pair, 500 toys are generated from

an angular distribution at that location in the parameter space, where each toy in the ensemble consists of n_k signal events fluctuated according to uncertainty in the fit to the B-mass distribution of the signal mode, corrected for the q^2 unfolded yield in $2 < q^2_{true} < 5$ GeV$^2/c^4$. For each signal toy a background toy is generated where the yield is fluctuated according to the background yields in the mass fits. The shape of the background toy is fluctuated according to uncertainty in the binned template. The original background template is subtracted from the background toy, leaving a residual background toy that is combined with the signal toy, and fit with the acceptance modified angular distribution. This approach ensures that the statistical uncertainty in both the signal yield, background yield, and background template is incorporated into the Felman Cousins likelihood surface. The likelihood ratio for each toy at (A_{FB}, F_H) is calculated as

$$\mathcal{R}_i = \frac{\mathcal{L}(x;(A_{FB},F_H))^{gen}_i}{\mathcal{L}(x;(A_{FB},F_H))^{best}_i}, \tag{4.13}$$

where $\mathcal{L}(x;(A_{FB},F_H))^{gen}_i)$ is the negative log-likelihood constructed out of the i^{th} dataset and the angular pdf evaluated at the sampled (A_{FB}, F_H) point denoted as $(A_{FB}, F_H)^{gen}_i$. Similarly, $\mathcal{L}(x;(A_{FB},F_H))^{best}_i)$ is the negative log-likelihood constructed out of the i^{th} dataset and the angular pdf evaluated at the best-fit point which resulted from the fit to this dataset. Two ratio terms are subsequently computed, $\mathcal{R}^{data}_i$ and $\mathcal{R}^{toy}_i$, where Eq. 4.13 is calculated for the master and toy datasets respectively. When evaluating the statistical uncertainty on toys the 'master dataset' is replaced with a master toy dataset, in order to evaluate the expected statistical uncertainty.

For each sampled (A_{FB}, F_H) point, the p_i-value is given by

$$p_i = \frac{N_{toys}(\mathcal{R}^{toy}_i > \mathcal{R}^{data}_i)}{N_{toys}}, \tag{4.14}$$

where N_{toys} is the number of toys generated at a particular point in the (A_{FB}, F_H) parameter space. The 1σ, 2σ, 3σ, contours are found by taking the constant contour in the (A_{FB}, F_H) space where the p_i values correspond to $< 68\%$, $< 95\%$ and $< 99.7\%$. It is worth noting that the statistical uncertainty from the yields and shapes of the background templates are intrinsically included within the Feldman Cousins intervals

using this residual background subtraction method in addition to the uncertainty in the signal yield and shape. Studies showing the impact of the individual background subtractions can be found in Sec. 4.12.

4.10.4 Control Mode Angular Fit

The expected angular distribution of $B^+ \to J/\Psi(\to e^+e^-)K^+$ decays has $A_{FB} = 0$ and $F_H = 0$ [49]. The angular fit to extract these observables was done using the fully selected data with a cut on the constrained mass of $m_{const} > 5185$. This selection ensures a very clean data sample, allowing a validation of the acceptance corrections and the signal-only fit method, without requiring background subtraction. The angular fit is performed as a single fit to the combined data for Run1 and Run2. The result of this fit is shown in Fig.4.29 and the statistical uncertainty evaluated in Fig. 4.30 where the confidence intervals have been computed using the Feldman Cousins method detailed in Sec. 4.10.3. The global value lies is in agreement with the SM prediction at the 1σ uncertainty level. The best fit result lies at $A_{FB} = 0.001$, and $F_H = 0.001$, and is compatible with the SM at 1σ.

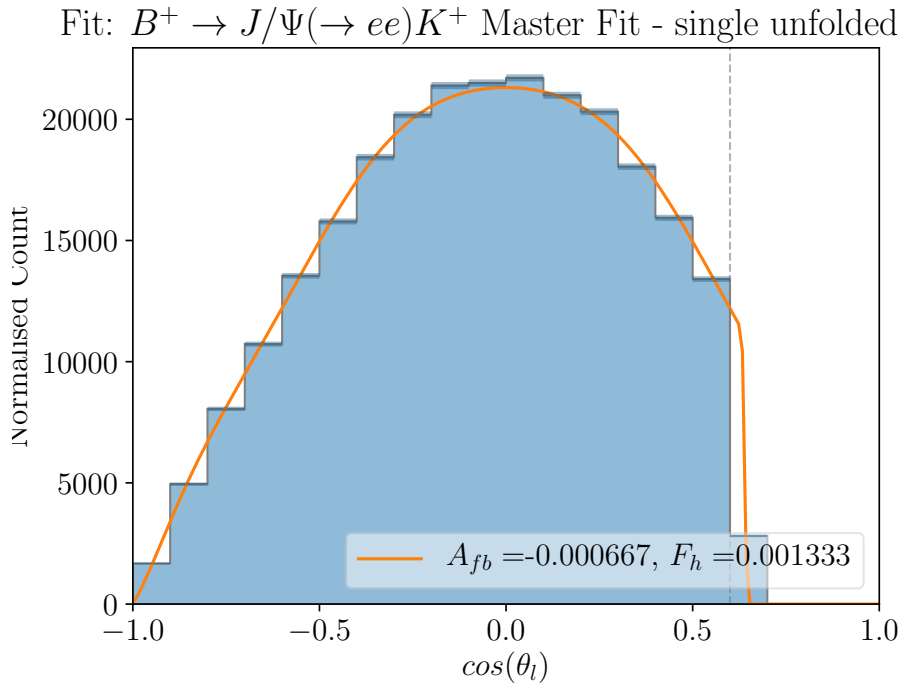

Figure 4.29: The global best fit to the angular parameters is $A_{FB} = 0.000667$, and $F_H = 0.00133$. The full result is given as a likelihood surface in Fig. 4.30.

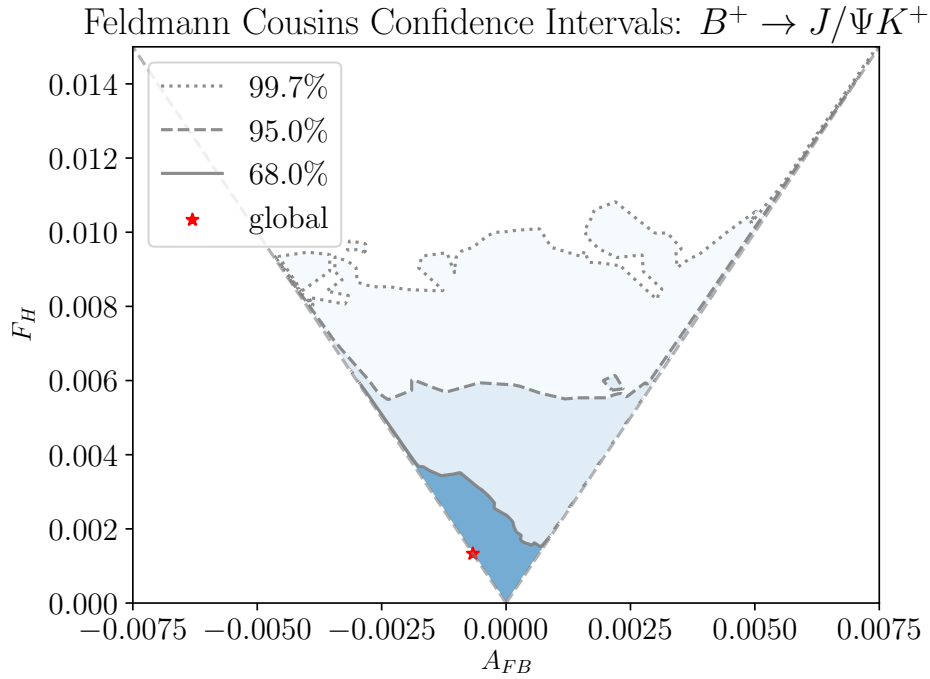

Figure 4.30: Feldman Cousins confidence interval on the global fit to the Run1 and Run2 $B^+ \to J/\Psi(e^+e^-)K^+)$ data. Each point in A_{FB}, F_H is evaluated with 500 toy datasets and a p-value calculated. The 1σ, 2σ and 3σ contours are shown. The global best fit to the angular parameters is $A_{FB} = 0.001$, and $F_H = 0.001$. The likelihood surface is compatible with the SM at less than 1σ.

4.10.5 Propagation to the rare mode

The applicability of the corrections to the simulation to the rare mode due to different q^2 and phase-space criteria between the control and the rare mode is studied by performing the angular fit to $B^+ \to J/\Psi(\to e^+e^-)K^+$ decays fully selected data in the most rare mode like region of $\text{ProbNN}_{\text{shell}} < 0.3$. Details of the $\text{ProbNN}_{\text{shell}}$ definition can be found in Sec 4.5.3.

By selecting control mode events that closely match the phase space of the rare mode, the validity of the corrections and their impact on the angular observables can be studied. The angular fit to this sub-dataset is performed using the same method as in Sec. 4.10.2 and the uncertainties estimated using the Feldmann Cousins method detailed in Sec. 4.10.3. The results of the fit and the statistical uncertainty can be found in Fig. 4.31, where the global best fit lies at $A_{FB} = -0.0002$, and $F_H = 0.0003$. The likelihood surface is compatible with the SM at less than 1σ. The result of this study suggests that the corrections applied have no significant dependence on q^2 and are compatible with the rare mode kinematics.

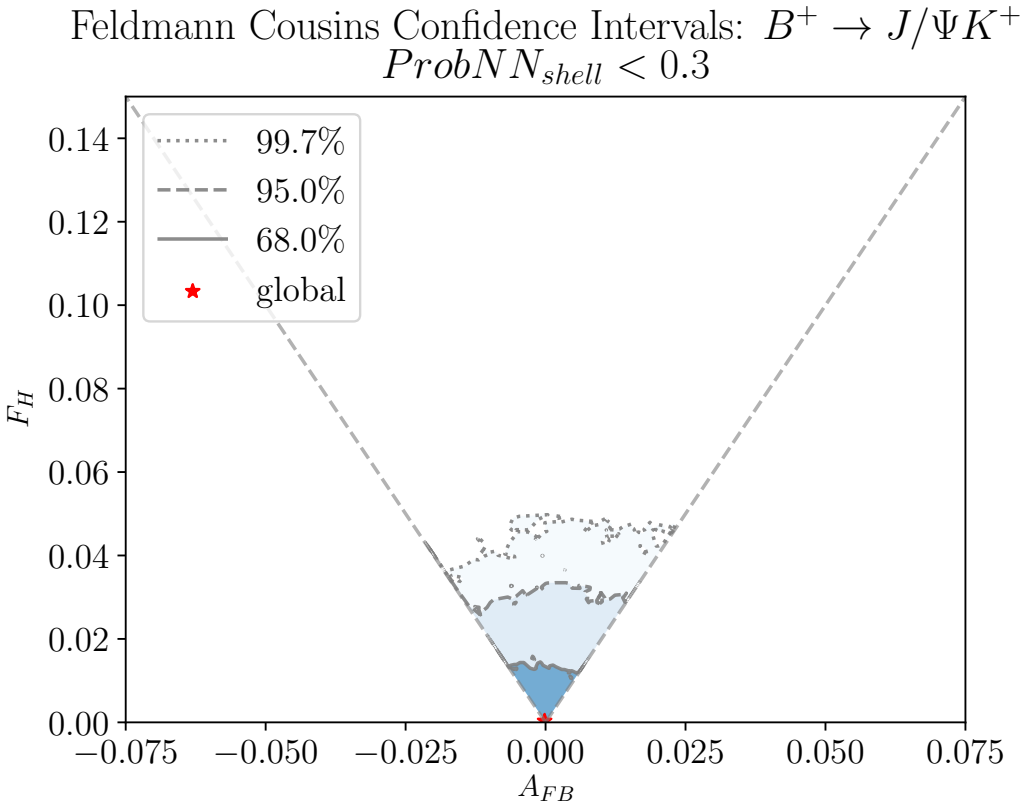

Figure 4.31: Feldmann Cousins confidence interval on the global fit to the entire Run1 and Run2 $B^+ \to J/\Psi(^+e^-)K^+$ data with $ProbNN_{shell} < 0.3$ selection criteria. Each point in A_{FB}, F_H is evaluated with 500 toy datasets and a p-value calculated. The 1σ, 2σ and 3σ contours are shown. The global best fit to the angular parameters is $A_{FB} = -0.0002$, and $F_H = 0.0003$.

4.10.6 Rare Mode Toy Studies

To investigate the sensitivity of the angular fit to the rare mode, a toy study was conducted. The master toy angular distribution of the signal was generated using $A_{FB} = 0.00$ and $F_H = 0.00$, modulated by the angular acceptance in each trigger category and run period[2]. To obtain uncertainties that realistically describe the rare mode data, the toy studies include a background subtraction component. The angular distributions of the backgrounds were generated according to the distributions given in Sec. 4.7.5. The number of signal and background candidates for the master toy were taken from the mass fit to the rare mode described in Sec. 4.7. The signal yield is corrected for the unfolding procedure, and the background yields are generated from a multivariate normal that describes the correlations between background yields. For each subsequent toy in the Feldman Cousins procedure, the yields are fluctuated according to the uncertainties detailed in Sec. 4.7. For each toy the six trigger categories and run periods are generated individually, then combined for the final angular fit.

The signal toy is generated from a q^2_{true} distribution, however the background components are not. As the background may not follow the same unfolding scheme as the signal, the backgrounds cannot be simply unfolded using the signal model and used to generate a true q^2 distribution, therefore, a residual background method was developed to study the impact of the background subtraction on the final fit. This involves generating a toy background from the reconstructed template, where statistical fluctuations in both the yield and shape are taken into account. The prior template is subtracted from the toy background to get a residual background component, and this residual is unfolded. This means that rather than relying on fully unfolding background distributions, which may be subject to variations that are not well modelled, the residual component is the only part that is unfolded. These residual events by definition contain only the background events not subtracted by the prior. As these events get unfolded with the signal it gives a realistic model of the residual background component in the angular fit. It should be

[2]The SM expectation lies very close to (0,0) [49]

noted that the residual angular distribution can take negative weights in each bin due to statistical fluctuations.

The angular fits to an example master toy dataset can be found in Fig. 4.32, generated consistent with yields taken from the rare mode Run1 and Run2 mass fits, found in Sec. 4.7. The resulting confidence intervals are shown in Fig. 4.33, with contour bands showing 1σ, 2σ, and 3σ statistical uncertainties. The likelihood surface of the toy Feldman Cousins is compatible with the generated value, and indicates the magnitude of the uncertainty intervals.

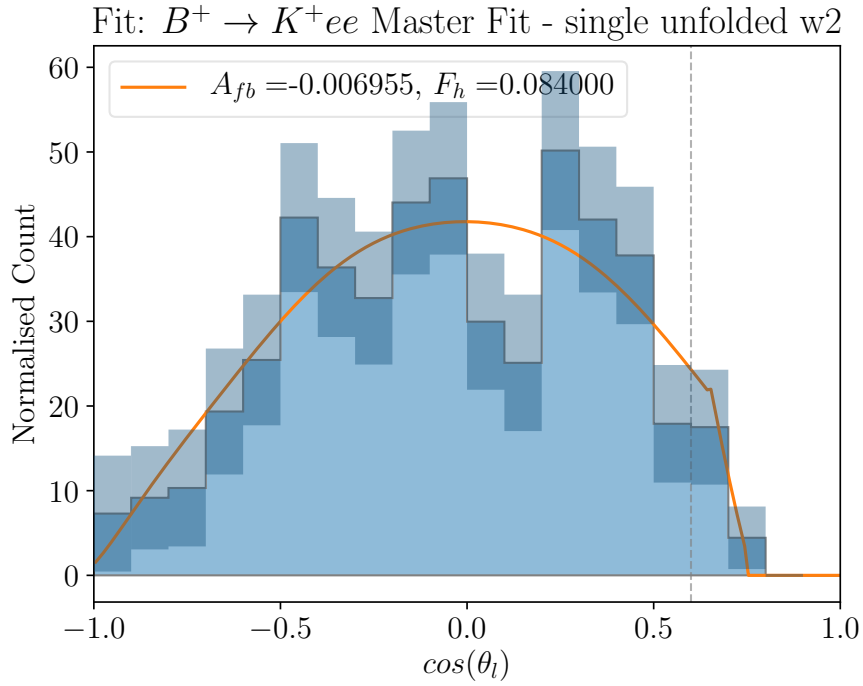

Figure 4.32: Example generated toy dataset of $B^+ \to K^+e^+e^-$decays at $F_H = 0$ and $A_{FB} = 0$, consistent with yields obtained from the mass fit. The shaded region on each bin represents the errorbar.

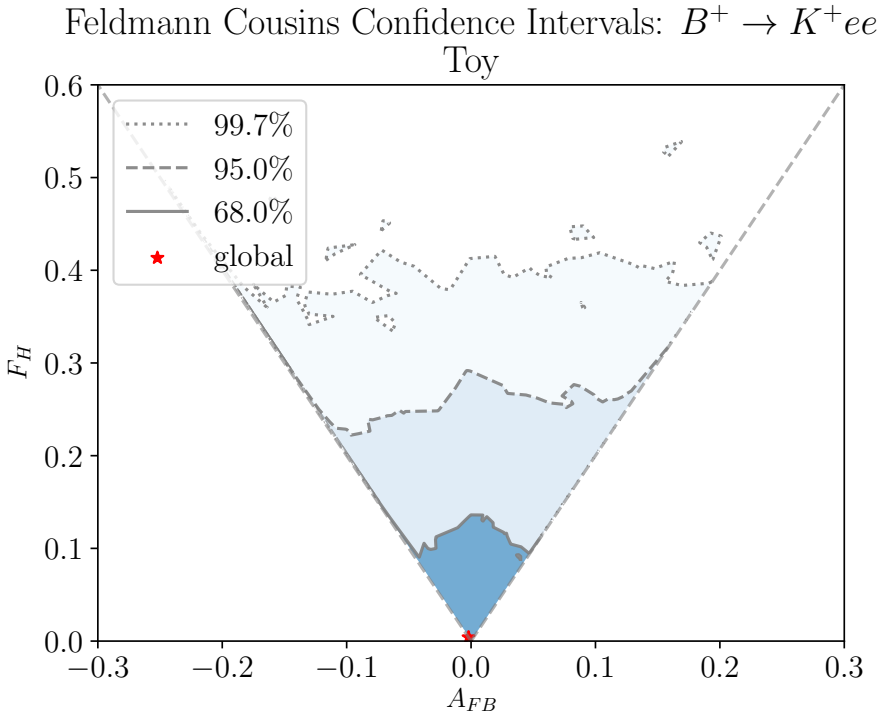

Figure 4.33: Feldman Cousins confidence interval on the global fit to the $B^+ \to K^+e^+e^-$ toy data, with toy background residual subtractions included. The global fit to the master toy dataset is shown with the red star, and the 1σ, 2σ, 3σ counters are shown. At each point in A_{FB} and F_H 500 toys are generated to calculate a p-value and produce the Feldman Cousins confidence contours.

4.11 Rare Mode Experimental Data Angular Fits

The angular fit to background subtracted $B^+ \to K^+e^+e^-$decays in a true q^2 bin ${q^2}_{true} \in [2, 5]$ GeV^2 is presented in this section. The data is initially selected as detailed in Table 4.9. The mass fits used to obtain yields for signal and background components are shown in Fig. 4.22, and the yields given in Tables. [4.10, 4.11]. A selection on the constrained mass of $m_{HOP} > 4900$ MeV is applied. The background templates used to subtract the background components are given in Fig. 4.23. The final background subtracted angular distribution is obtained in the narrower B-mass window of $5000 < m_{Kee} < 5600$ MeV to reduce total background subtracted while maintaining a high signal yield efficiency. This results in a background efficiency of 53% for the combinatorial and 65% for the partially reconstructed, at an efficiency of 90.6% for the signal events.

The angular fit to the data follows the same procedure as in the control mode in Sec. 4.10.4, and the two dimensional confidence intervals are evaluated using the same procedure presented in Sec. 4.10.6. The angular fit to the data can be found in Fig. 4.34, where the best fit angular observables are found to be $A_{FB} = 0.036$, and $F_H = 0.072$. The Feldman Cousins confidence intervals given in Fig. 4.35 show that the likelihood surface is in agreement with the standard model within a 1σ uncertainty. The datasets from Run1 and Run2 have been fit independently, are are found to be compatible.

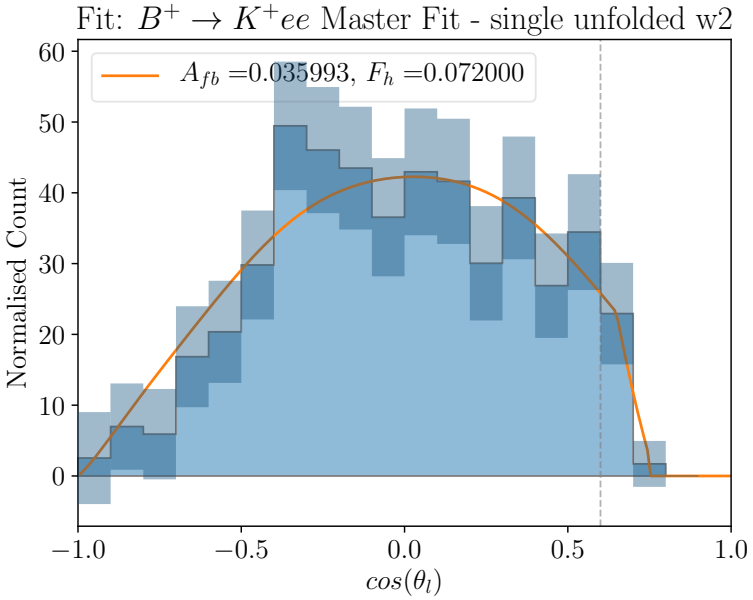

Figure 4.34: Angular fit to the total rare mode data, in $2 < q^2_{\text{TRUE}} < 5$ GeV$^2/c^4$, with backgrounds subtracted. The global best fit to the angular parameters is $A_{FB} = 0.036$, and $F_H = 0.072$. The full result is given as a likelihood surface in Fig. 4.35. The shaded region on each bin represents the errorbar.

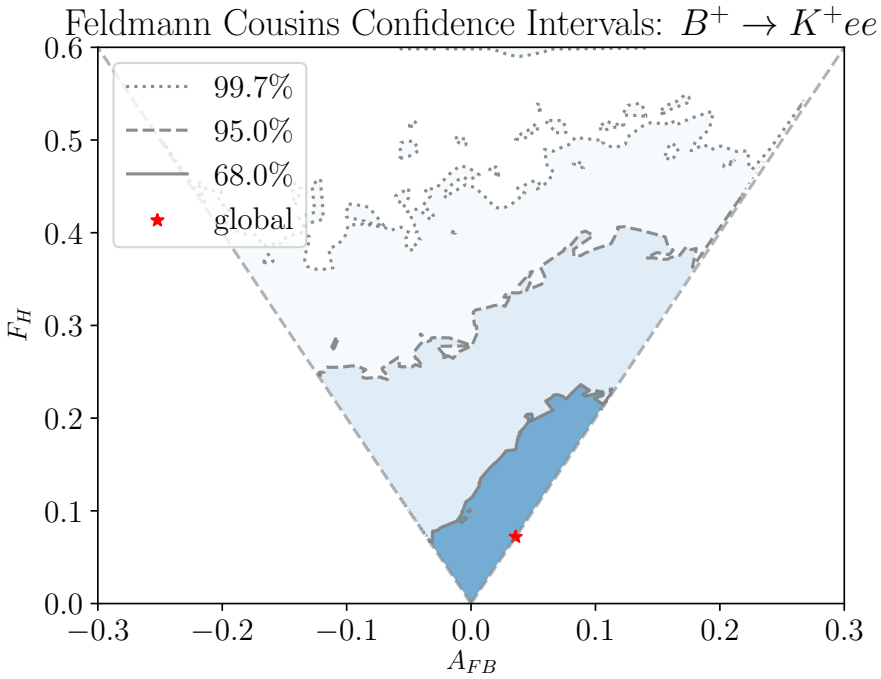

Figure 4.35: Feldman Cousins confidence interval on the global fit to the $B^+ \to K^+e^+e^-$ experimental data, with backgrounds subtracted. The global fit to the master dataset is shown with the red star, and the 1σ, 2σ, 3σ counters are shown. At each point in A_{FB} and F_H 500 toys are generated to calculate a p-value and produce the Feldman Cousins confidence contours. The expected SM value at the origin is comfortably contained within the 1σ confidence interval.

4.12 Systematic uncertainties

The following section details the studies undertaken to evaluate the systematic uncertainties associated with the angular analysis presented in this chapter. The key areas addressed are the uncertainty from limited sample size of the simulation in Sec. 4.12.1 and the uncertainty contributed from yields and limited stats in templates to the background subtraction in Sec. 4.12.2. However, the background template shape uncertainty and background yield uncertainty is already included in the Feldman Cousins confidence intervals given in Sec. 4.10. Other systematic uncertainties are evaluated, and found to be small, details of these studies are included in the following section. Additional systematics for the unfolding method and mass fit line shapes are given. All systematics are assumed to be uncorrelated.

4.12.1 Systematic uncertainties from the acceptance correction

The angular fit relies on the simulation to precisely model the acceptance correction. The simulation has been extensively corrected using data from the control mode, however there is still an uncertainty associated with the finite statistics of the simulation when modelling the acceptance correction. There are two sources of uncertainties related with the acceptance correction. Firstly the statistical uncertainty of the simulated sample itself, and secondly the choice of model used to describe the acceptance correction.

The statistical uncertainty of the acceptance correction was estimated by generating a signal toy with a yield 1000 times the expected signal yield at the SM value, using the nominal acceptance correction. Different acceptance corrections are subsequently derived by bootstrapping the simulated sample and repeating the fit to the nominal dataset, but with a modified acceptance correction. A density map showing the spread of the resulting fits in the two dimensional A_{FB}, F_H plane corresponds to the statistical uncertainty of the acceptance correction. This spread is shown in Fig. 4.36 and is significantly below the statistical uncertainty in the rare mode data.

The impact of the choice of order used in the Legendre polynomial was studied as well,

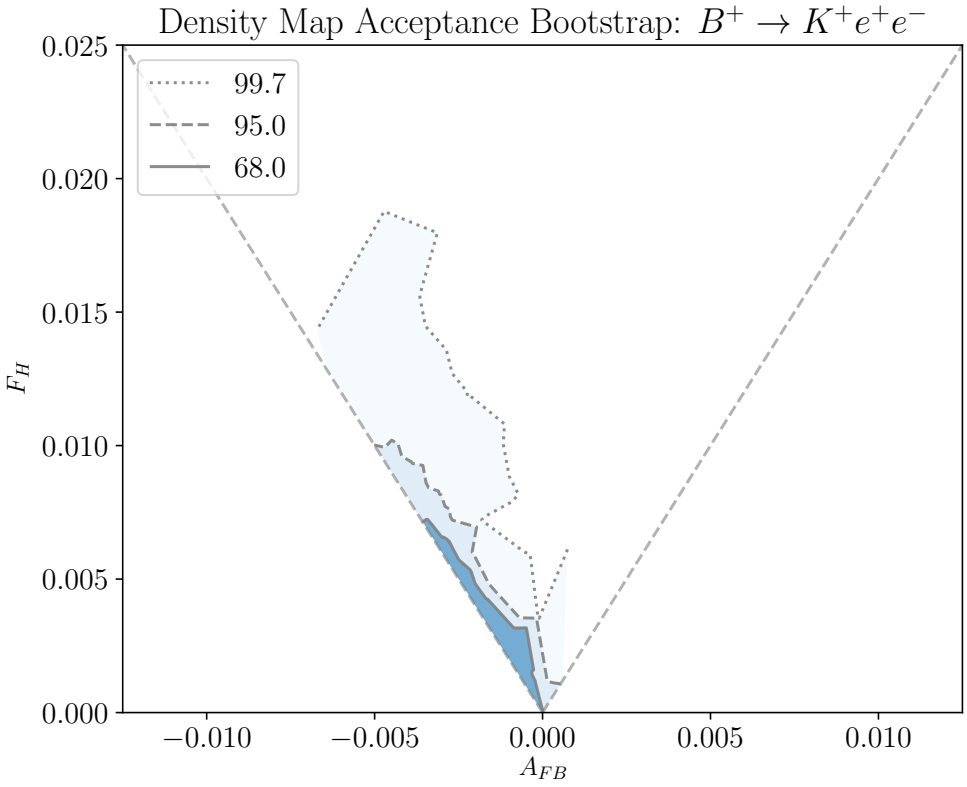

Figure 4.36: Density plot of the spread of best fit results when bootstrapping the MC to evaluate the impact of limited statistics on the acceptance correction. Contours show the relative density within 68%, 95% and 99.7% of the peak density.

where a sample was generated with a polynomial 4 orders higher for each trigger category and fitted with both the nominal orders and the higher orders. The discrepancy between the two fits corresponds to the systematic uncertainty due to the choice of the polynomial order. This was found to have a negligible impact on the fit result when compared to the statistical uncertainty in the rare mode data sample.

4.12.2 Systematic uncertainty due to the background subtraction

In this section the systematic uncertainties related to the background subtraction are evaluated. The uncertainties related to the statistical uncertainty of the angular templates as well as the background yields are already accounted for in the FC contours of Sec. 4.10.6.

4.12.2.1 Combinatorial Background Subtraction

The combinatorial background subtraction takes the shape from the upper mass side band (UMSB) and subtracts this shape according to the yield from the final mass fit, as in section 4.7.5. This introduces two types of uncertainty, firstly the uncertainty from the yield of the fit, and secondly from the shape of the combinatorial estimate as taken from the UMSB due to a finite number of entries per bin. The first can easily be varied by generating toy background samples with yields taken from a normal distribution, with a central value coming from the fit yield and with a width from the uncertainty of the fit yield. The second uncertainty requires modulating the shape of the UMSB background template based on the Poisson uncertainty on the yield in each bin. A residual background component is then calculated as the subtraction of the toy background from the true background sample from the UMSB, appropriately scaled to the expected background yield from the mass fits. Both of these variations are included in the Feldman Cousins method, and are evaluated independently here. To evaluate the impact of this uncertainty on the angular fit in a true q^2 region the residual background was unfolded and fitted with a single master signal toy generated at the standard model value. The density of the spread of the fit results following this method gives an estimate of the uncertainty expected from subtracting a background template from the signal and fitting an angular distribution in a true $2 < q^2_{true} < 5$ GeV$^2/c^4$ bin assuming a signal only sample. The uncertainty estimation for the combinatorial subtraction can be seen in Fig. 4.37 where the contours represent the density of fit results. The contours are shown for 68%, 95% and 99.7% decrease in density from the peak density. The statistical uncertainty of the

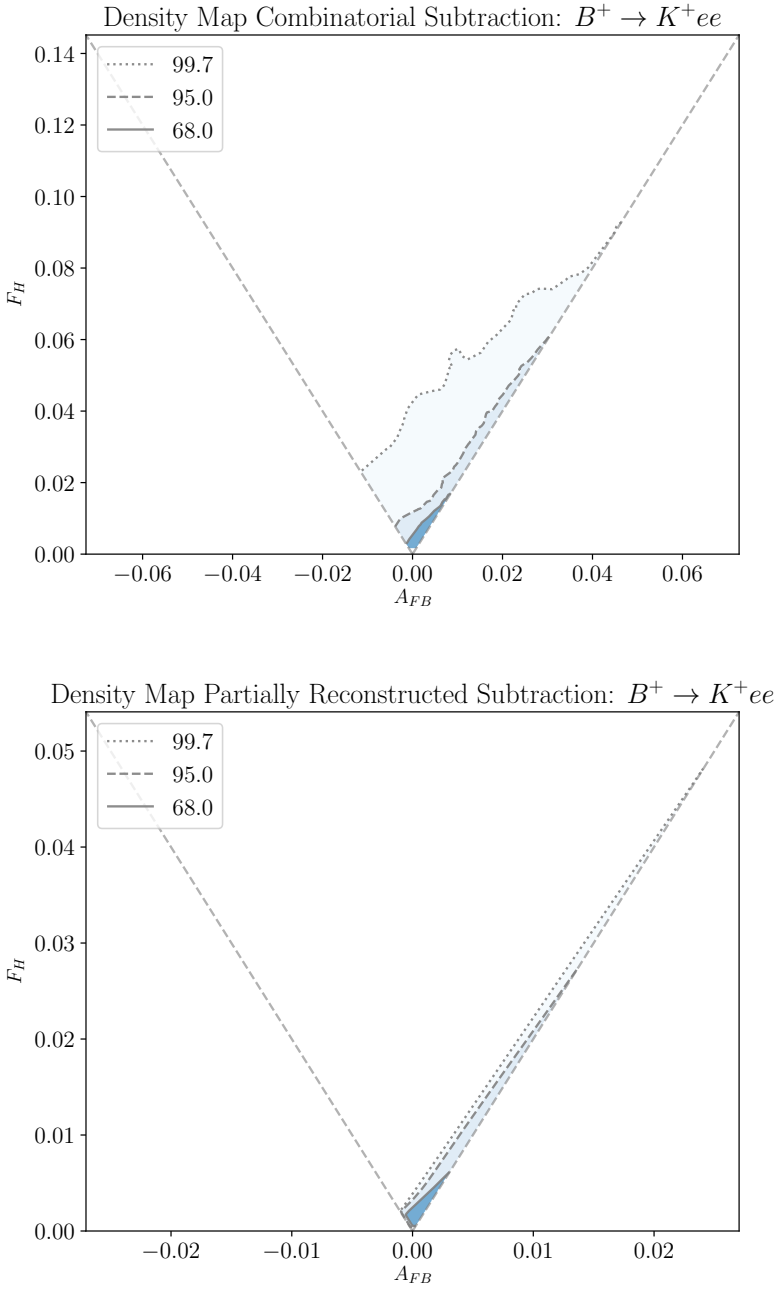

Figure 4.37: Density plot of the spread of best fit results when subtracting combinatorial background toys (top) and the partially reconstructed background (bottom). Contours show the relative density within 68%, 95% and 99.7% of the peak density.

background subtraction procedure is found to be small and is already included in the Feldman Cousins interval of the angular fit.

4.12.2.2 Combinatorial Background Subtraction Template Shape

In the previous section, Sec.4.12.2.1, the background subtraction is conducted using the template for the combinatorial background from the upper mass side band of the $B^+ \to K^+e^+e^-$ data. This is necessary because the angular shape of the combinatorial background cannot be obtained from the signal region, however, a systematic is needed to account for the displacement of the UMSB from the signal region. To study this the $B^+ \to K^+e\mu$ data sample is used, and the same background subtraction method is performed. However rather than using the same nominal fluctuated sample to generate toys and to subtract, the upper mass side band is fluctuated to generate toys, and the nominal background subtracted is from the signal region in m_{Kee}. This allows two templates from data in different mass ranges to be subtracted from each other. The upper mass side band is defined as $m_{Kee} > 5475$ MeV, and the signal region template defined as $m_{Kee} < 5350$ MeV. In order to factorise the systematic effect from the statistical effect the combinatorial BDT selection is not applied to the $B^+ \to K^+e\mu$ upper mass side band, this allows the uncertainty in the bins to be relatively small compared to the choice of template. The result shown in Fig.4.38 shows the contours of a density plot containing 68% and 95% of the best fit locations, however, comparison to Fig.4.37 should take into consideration that this study contains the uncertainty of both the background yield from the mass fit to the rare mode, and the statistical uncertainty from the sample size and shape of the background template. From this it can be concluded that the statistical uncertainty dominates any potential bias due to the template being taken from the upper mass side band.

4.12.2.3 Partially-Reconstructed Background Subtraction

The partially-reconstructed background subtraction uncertainty is estimated in the same manner as the combinatorial background in Sec. 4.12.2.1. The only difference being the template for the background is taken from the $B^{0,+} \to K^{*0,+}e^+e^-$ simulation sample. This resulted in a negligible uncertainty compared to the statistical uncertainty in the

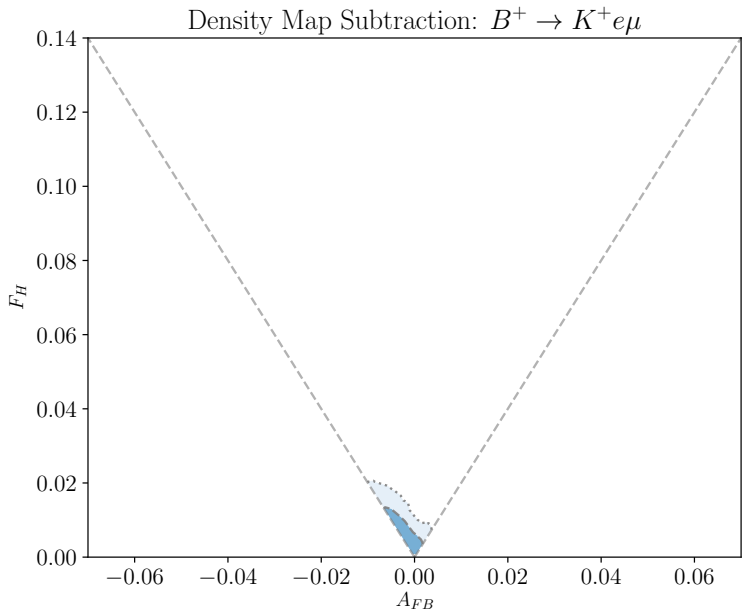

Figure 4.38: Density plot of the spread of best fit results when subtracting combinatorial background toys using $B^+ \to K^+e\mu$ data samples, where the background is generated using a template from the upper mass side band ($m_{Kee} > 5475$ MeV) and subtracted using a template from the signal region ($m_{Kee} < 5350$ MeV). Contours show the relative density within 68%, 95% and 99.7% of the peak density.

rare mode and is already included in the Feldman Cousins intervals of the angular fit.

The result can be seen in Fig.4.37.

4.12.3 Angular Systematics

Other systematics evaluated that influence the angular distribution are the systematic uncertainties associated with the correction weights; the mass independence of the angular distribution; the impact that the resolution of $\cos\theta_\ell$ has on the binned fit; and the q^2 dependence of the angular distribution across the q^2 bin $q^2 \in [2, 5]$ GeV^2.

4.12.3.1 Weight Systematics and `nSPDHits`

The corrections applied to the simulation samples have an associated uncertainty. The impact of this uncertainty on the angular distribution is studied in the simulation samples by comparing the distributions with all corrections applied and without any corrections applied. The distributions are split in six regions for the two data taking periods and the three trigger categories, and can be found in Fig 4.39 and Fig 4.40. The pull distributions below each plot show that the impact of the weights on the angular distribution is small, and given that the rare mode data has on the order of 100 times fewer events than the simulation samples, the statistical uncertainty dominates the uncertainty from the weight corrections. To study the potential bias caused by the correction weights to simulation samples a high statistics toy is generated at the SM value with the nominal weights applied to the acceptance correction. The high statistics toy is the fit using an acceptance correction where the weights are not applied. This demonstrates a reasonable approximation of the maximum bias that would be introduced by the weights. The impact of omitting the corrections to the simulation in rare-mode toys is shown in the left plot of Fig.4.41, where the 1σ contour of the density plot comfortably contains the generated point, and no bias is found.

An additional systematic uncertainty comes from the decision to exclude the multiplicity variable, nSPDHits, from the re-weighting algorithm used in this analysis. Motivation for this exclusion is detailed in the R_K analysis note [36], where it was found that re-weighting this variable lead to worse agreement in the bremsstrahlung fractions between data and

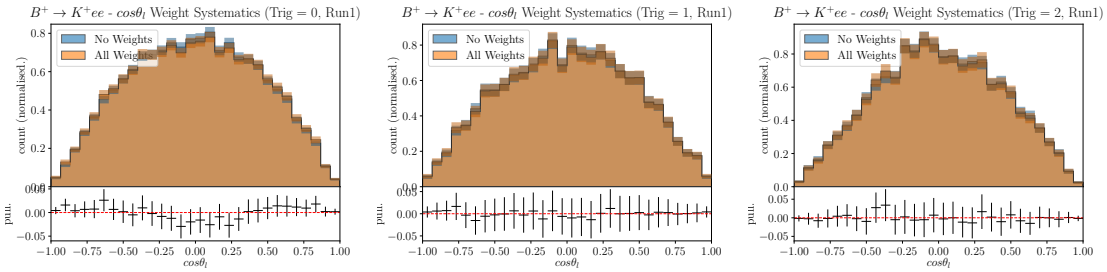

Figure 4.39: The impact of the corrections weights on the angular distribution is shown for the Run1 simulation samples in three trigger categories: eTOS (left); hTOS (center); TIS (right). In blue no weights are applied, and in orange the full final weights.

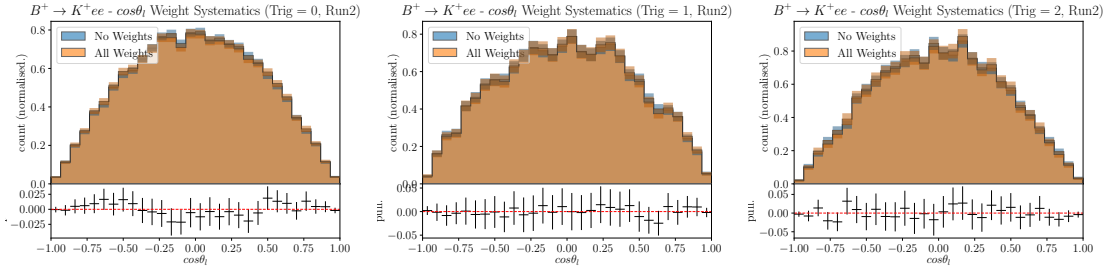

Figure 4.40: The impact of the corrections weights on the angular distribution is shown for the Run2 simulation samples in three trigger categories: eTOS (left); hTOS (centre); TIS (right). In blue no weights are applied, and in orange the full final weights.

simulation. However, the multiplicity contains important information that is related to the reconstruction of electron kinematic variables, so a systematic uncertainty is estimated to account for not including this in the re-weighting. A correction weight for `nSPDHits`, referred to as w_{nSPD}, is modelled using the control mode fully selected data and simulation samples, with a cut applied on the constrained mass of $m_{const} > 5185$ MeV, and this model used in the prediction of weights for the rare mode simulated samples. A similar method detailed above is used to model this systematic uncertainty. A large toy sample is generated with the nominal weights as before, however when fitting rather than using an acceptance correction generated using no weights, an acceptance correction is calculated using the original correction multiplied by w_{nSPD}.

In these studies if there is a bias associated with not applying the weights, or not correcting for nSPDHits, it will manifest as a small fixed shift away from the A_{FB}, F_H values where the toy was generated. The results from these studies can be seen in the right plot of Fig.4.41, where the contours show the region containing 68%, 95% and 99.7%

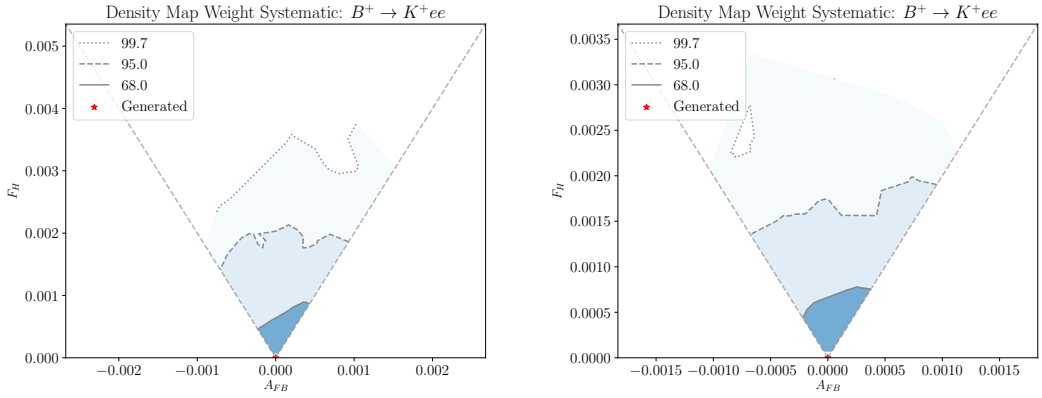

Figure 4.41: Bias study using high statistic toys with and without correction weights applied. The impact of fitting a high statistics toy generated with the nominal weight used to compute the acceptance correction, then fitted with an acceptance correction computed using no weights is shown on the left. The impact of nSPDHits not being re-weighted is shown on the right, where the large stats toy is generated with the nominal weights used in the computation of the acceptance correction, then fitted using an acceptance correction where nSPDHits has been re-weighted as well. Contours show the relative density within 68%, 95% and 99.7% of the peak density. (Note: The 99.7 contour shape is a result of limited statistics in the toy study.)

of the toy samples, and the left plot shows the impact of not applying any weights, and the right the impact of re-weighting nSPDHits. Therefore for an SM like result, there is no notable bias. A further element of the study was performed looking at the impact of a potential bias for A_{FB} and F_H values further away from the SM, and it can be seen in Fig.4.42 that no bias is introduced for non SM like results.

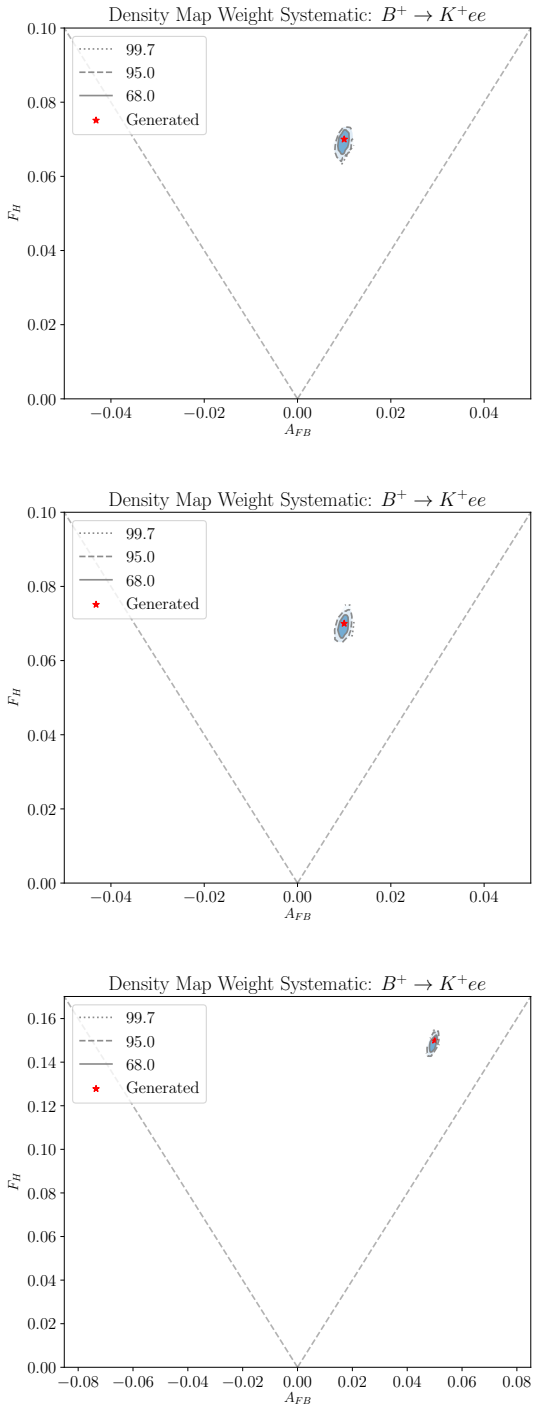

Figure 4.42: Extension of the results shown in Fig. 4.41, to study the impact of non-SM like values of A_{FB} and F_H on the uncertainty. Impact of not applying the weights (top), weighting nSPDHits (centre), weighting nSPDHits at a more extreme pair of values (bottom). In all cases the spread of the contours represents the statistics used in the toy generation, and any bias would be seen as a shift away from the generated point. Contours show the relative density within 68%, 95% and 99.7% of the peak density.

4.12.3.2 Mass Independence of angular distribution - combinatorial background

The template of the angular distribution for the combinatorial background comes from the upper mass side band in the data. It is not expected that the angular distribution should change substantially when extrapolating from the upper mass side band to the signal region, however, the selection process could introduce a bias. Such a bias is tested for using the $B^+ \to K^+e\mu$ sample as the combinatorial background proxy. Two mass regions are chosen, above $m_{Kee} > 5475\,\text{MeV}$ and below $m_{Kee} < 5350\,\text{MeV}$ and the angular distributions compared in Fig.4.43. No significant deviations are seen to suggest that the upper mass side band cannot be used to provide the template for the combinatorial background angular distribution. An additional check was made using simulation data for the control mode like events of the type $B^+ \to J/\Psi(e^+e^-)X$, where X is any hadron, and a background category is chosen such that the event is considered combinatorial. The angular distribution is compatible between the upper mass side band and signal region for this sample. The $B^+ \to J/\Psi(e^+e^-)X$ sample has a significantly different angular shape to the $B^+ \to K^+e^+e^-$ combinatorial so a further check was made using the rare mode simulation samples, also selected to be of combinatorial background category. This check was limited in power due to simulation sample size, however was in agreement with the shape of combinatorial found in rare mode signal events from the upper mass side band. These results are also shown in Fig. 4.43. The angular distribution is considered to be independent of m_{Kee}.

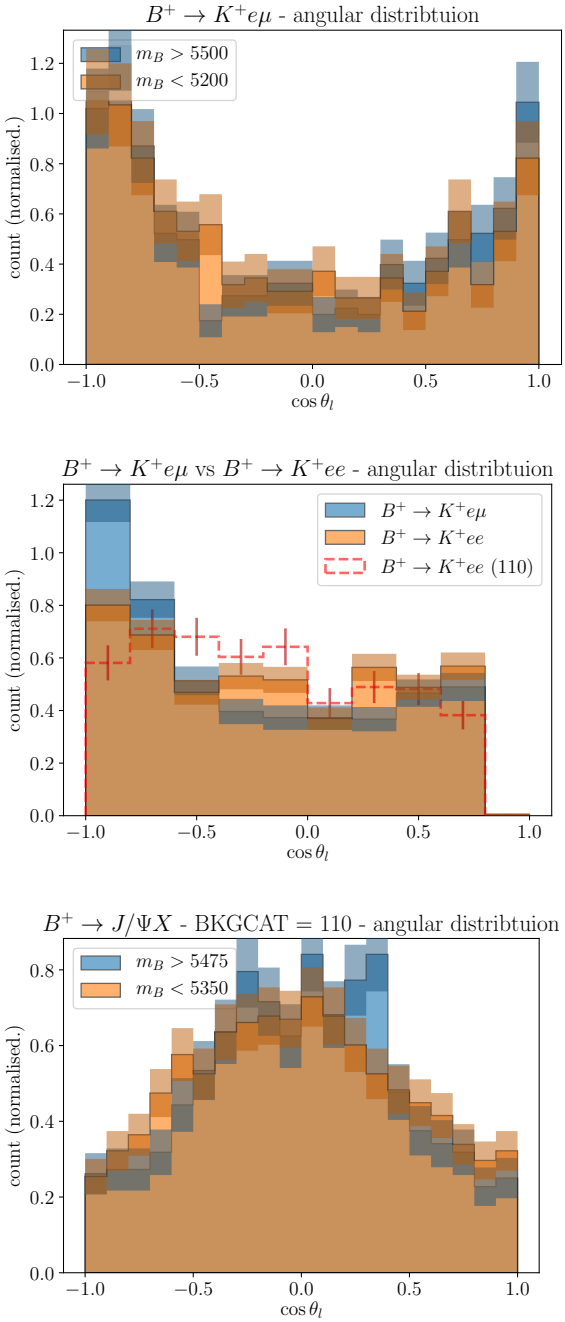

Figure 4.43: The angular distribution of $B^+ \to K^+e^\pm\mu^\mp$ shown for the upper mass side band ($m_{Kee} > 5475\,\text{MeV}$) and the signal region ($m_{Kee} < 5350\,\text{MeV}$) (top). The regions are separated to ensure unique distributions, with a buffer between. The angular distribution of the $B^+ \to Ke^\pm\mu^\mp$ sample in blue compared to the upper mass side band of the rare data in orange, and combinatorial matched simulation containing at least one signal decay for the rare mode simulation in red (centre). These samples have the veto applied. The angular distribution for the upper mass side band and signal region for $B^+ \to J/\Psi X$ simulation samples, selecting combinatorial events (bottom).

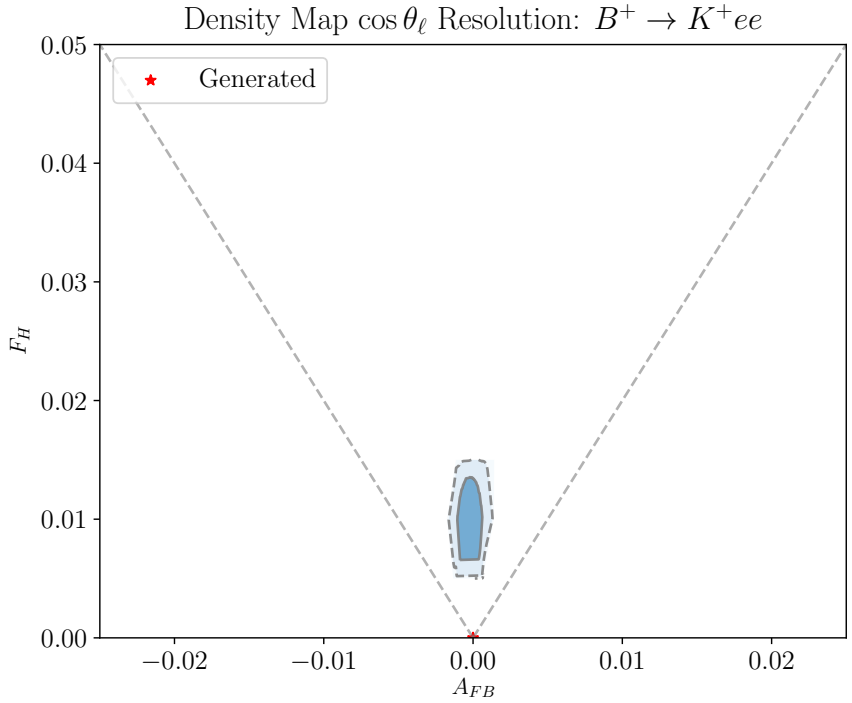

Figure 4.44: Systematic uncertainty associated with the resolution of $\cos\theta_\ell$, shown on an F_H scale representative of the statistical uncertainty in the rare mode. Total bias on the order of 0.01 in F_H.

4.12.3.3 $\cos\theta_\ell$ Resolution Systematic

The resolution in $\cos\theta_\ell$ is not unfolded in the angular fitting procedure, therefore a systematic needs to be assigned to the impact of migration of true $\cos\theta_\ell$ events to neighbouring reconstructed bins. This systematic uncertainty is evaluated by generating high-statistics toys according to the SM, smearing these generated events according to the resolution as a function of $\cos\theta_\ell$, applying the acceptance correction using an accept-reject methodology, and fitting the angular distribution. The result of this study can be seen in Fig. 4.44 where the shift away from $A_{FB} = 0$, $F_H = 0$ corresponds to the bias due to not accounting for the resolution. This bias in A_{FB} is negligible, and in F_H corresponds to $\sim 4\%$ of the statistical uncertainty of the final result. For the purpose of this thesis this bias is ignored. However, in the final publication this uncertainty will be re-estimated by generating toys from the best-fit point of the final fit, reassessing the bias and correcting the central fit point by this value.

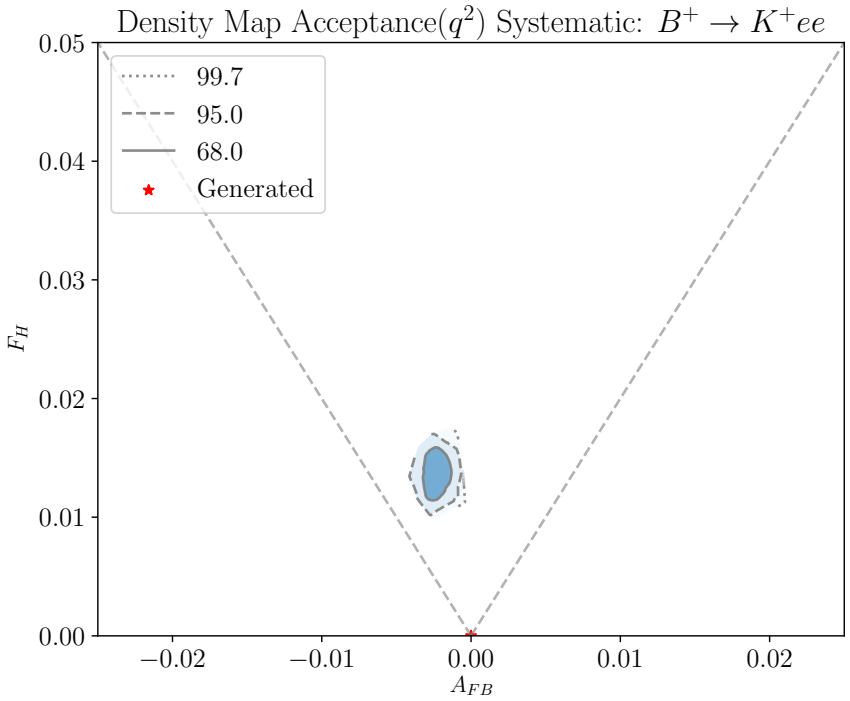

Figure 4.45: Systematic uncertainty associated with the dependence of q^2 on the acceptance correction. Total bias on the order of 0.01 in F_H.

4.12.3.4 q^2 Dependence Systematic

To investigate the q^2 dependence of the acceptance correction a q^2 range $3.5 < q^2 < 5$ was used to obtain the correction, that was fitted to the q^2 range $2.0 < q^2 < 3.5$. The results of fitting high statistic toys is shown in Fig. 4.45. This bias in A_{FB} is negligible, and in F_H corresponds to $\sim 6\%$ of the statistical uncertainty of the final result. For the purpose of this thesis this bias is ignored. However, in the final publication this uncertainty will be re-estimated by generating toys from the best-fit point of the final fit, reassessing the bias and correcting the central fit point by this value.

4.12.4 Unfolding Systematic

The nominal unfolding method using in Sec. 4.8 is a naive matrix inversion method. The choice of unfolding method could introduce a source of systematic uncertainty into the final angular fit, so an alternative method based on a Bayesian approach was compared [50].

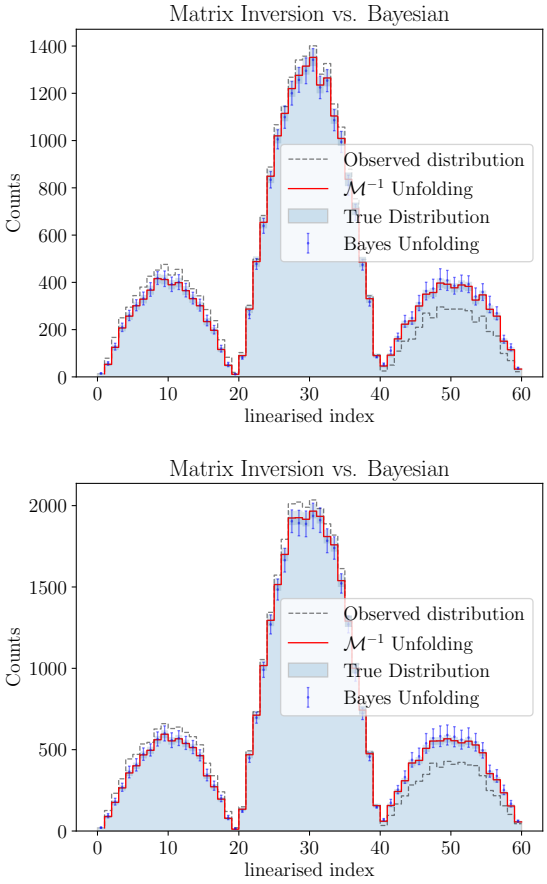

Figure 4.46: Comparison of naive matrix inversion and Bayesian unfolding methods. Run1 (top), Run2 (bottom). The observed reconstructed distribution is shown by the grey dashed line, and the true distribution by the light blue histogram. The Matrix unfolding method is shown by the red line and the Bayesian unfolding method shown by the blue points with error bars.

In Fig. 4.46, the observed and true distribution of linearised index for $B^+ \to K^+e^+e^-$ simulated samples are shown by the grey dashed line and blue histogram respectively, where the index is used to describe the transition between three regions in q^2 while retaining the $\cos\theta_l$ distribution. More details can be found in Sec. 4.8. In Fig. 4.46 the matrix inversion method is shown in red to be in perfect agreement with the true distribution, the Bayesian unfolding method is shown by the blue points with error bars is in agreement with the matrix unfolding method. The differences between the two methods is small and the choice of unfolding method will therefore introduce a negligible systematic uncertainty.

Type	Nominal central value	Nominal Fit σ	Bootstrap σ
(Run1) Signal	341	41	2.6
(Run1) Combinatorial	171	62	2.6
(Run1) Part-Reco	185	29	0.9
(Run2) Signal	428	35	2.5
(Run2) Combinatorial	173	55	2.5
(Run2) Part-Reco	226	45	1.8

Table 4.12: Comparison of the nominal mass fit results to the variation observed in the signal line shape bootstrap study. The spread of yields due to bootstrapping is significantly smaller than the uncertainty in the mass fit yields.

4.12.5 Mass fitting line-shape systematic

The rare mode simulated samples are used to obtain the mass-fit line-shape templates used in the fits in Sec. 4.7 and shown in Appendix. A.2.2. The limited MC statistics introduce an uncertainty into the angular fit through the impact of the line-shape on the background yields; which are used to subtract an appropriate level of background from the rare mode data sample. To investigate the impact of the line-shape uncertainty the simulated samples were bootstrapped, re-fit, and used in the rare mode mass fit on the nominal data. This step was repeated 300 times and the spread of yields recorded. As can be seen in Table. 4.12 the spread observed in the yields as a result of re-sampling and re-fitting the templates is significantly smaller than the uncertainty obtained in the nominal fit itself and is as such ignored from the list of systematic uncertainties. The central value of the yields in the study matches each of the nominal yields.

4.13 Other Cross Checks

A number of cross checks have been performed to validate that the data is understood. The cross checks presented here include: the L0 trigger performance; the control mode Bremsstrahlung category modelling; the modelling of m_{HOP}; studies into the composition of partially reconstructed backgrounds; multiplicity studies; multiple candidates studies. This list is not exhaustive, other checks have been performed but the most important cross checks are presented for clarity.

4.13.1 L0 Trigger performance

To validate the L0 trigger modelling in the simulation the control mode samples are used. The samples are fully selected with the mass constraint selection applied, additionally the simulated samples have all the correction weights applied. The mis-modelling of the relative fraction of events in each L0 trigger category is observed to be small, on the level of a few percent, indicating no significant bias. The results can be seen in Fig.4.47. These small differences in trigger efficiency could have an impact on the rare mode mass fit yields as the trigger ratios are taken from simulation. To explore if this would correspond to a significant systematic effect the percentage mis-modelling in each category was taken from Fig.4.47 and used to correct the rare mode trigger fractions in the mass fit. These differences result in a change to the yields as shown in Table. 4.13 and can be seen to be very small. Furthermore, the modelling of the triggers was explored using the rare mode proxy variable $ProbNN_{\text{shell}}$. The relative ratio of trigger categories is clearly different to the nominal selection, and some slight discrepancies are seen in how well the simulation models the data data in Fig 4.48. This is not considered to be significant based on the scale of variation seen in Table. 4.13.

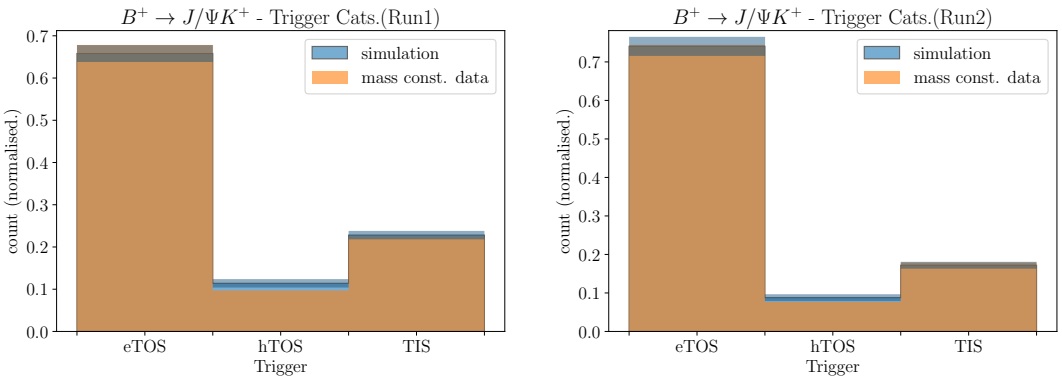

Figure 4.47: L0 Ratio plots for control mode fully selected sampes with the selection placed on the constrained mass.

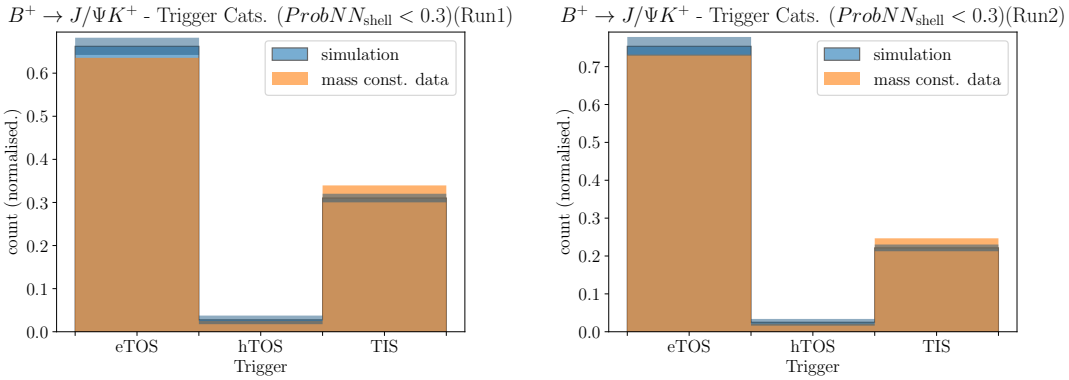

Figure 4.48: L0 Ratio plots for fully selected control mode, with the selection placed on the constrained mass, and with a $ProbNN_{shell} < 0.3$ selection placed to observe the trigger mixture in the *"rare mode"* like sample of the control mode data. The differences between Data and MC are applied to the rare mode MC, they have an effect much smaller than the statistical uncertainty of the signal and background determination from the mass fit.

	Nominal	Corrected
Run1		
Signal	348 ± 30	348 ± 26
Part-reco	188 ± 24	189 ± 23
eTOS Combinatorial	23 ± 16	23 ± 9
hTOS Combinatorial	99 ± 15	99 ± 14
TIS Combinatorial	38 ± 12	38 ± 11
Run2		
Signal	424 ± 35	415 ± 34
Part-Reco	223 ± 44	222 ± 45
eTOS Combinatorial	96 ± 27	90 ± 28
hTOS Combinatorial	34 ± 13	48 ± 13
TIS Combinatorial	49 ± 14	48 ± 14

Table 4.13: Yields from mass fits to Run1 and Run2 rare mode data, comparing the nominal trigger fraction mixture and the trigger fraction mixture corrected using Fig.4.48. All floating yields in the fit remain largely unchanged.

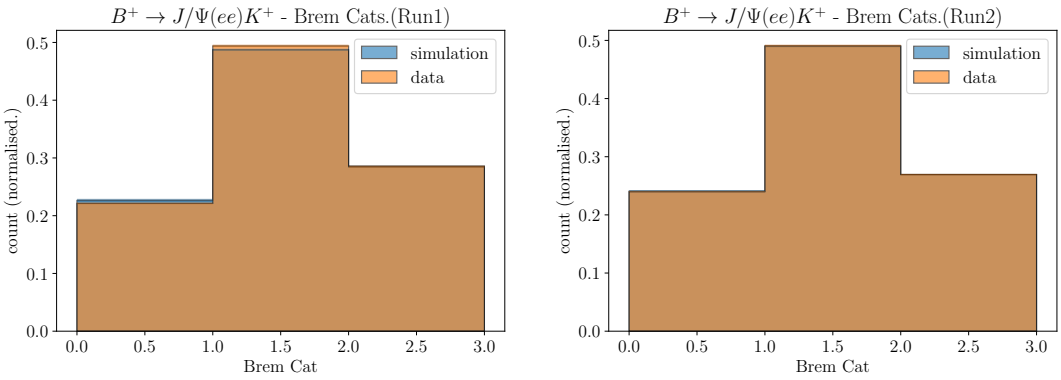

Figure 4.49: Bremsstrahlung Ratio plots for fully selected control mode with the selection placed on the constrained mass.

	γ 0	γ 1	γ 2
Run1 MC	0.226 ± 0.003	0.487 ± 0.002	0.286 ± 0.001
Run1 Data	0.221 ± 0.002	0.494 ± 0.002	0.285 ± 0.002
Run2 MC	0.240 ± 0.002	0.490 ± 0.001	0.270 ± 0.001
Run2 Data	0.240 ± 0.001	0.491 ± 0.001	0.269 ± 0.001

Table 4.14: Control mode bremsstrahlung fractions, as visualised in Fig. 4.49.

4.13.2 Control mode bremsstrahlung modelling

The fractions of bremsstrahlung categories in the control mode are shown in Fig.4.49 and can be seen to be in very good agreement. This gives confidence that the fractions of each category can be taken from simulation and fixed in the mass fit given the statistics of the rare mode. The numerical values are given in Table. 4.14.

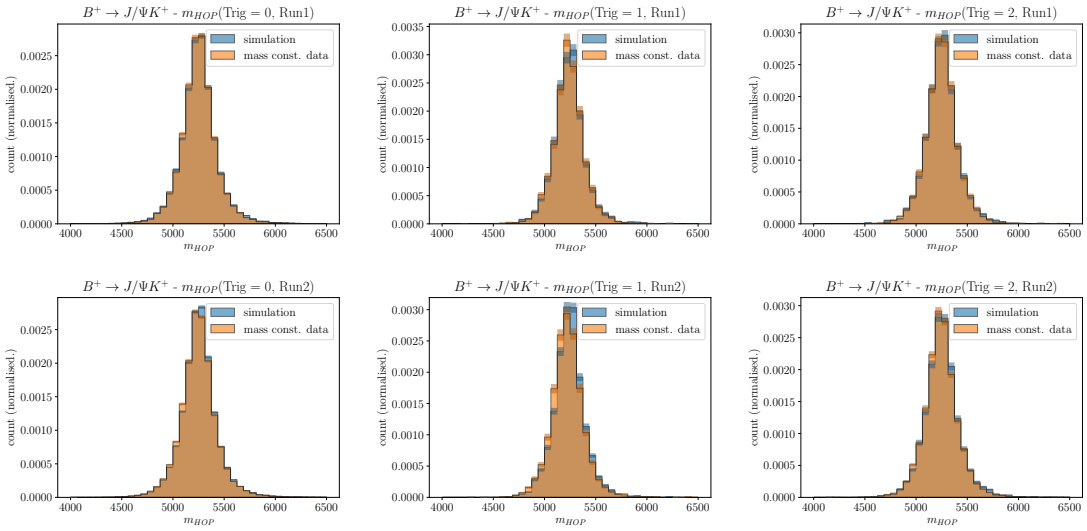

Figure 4.50: HOP mass modelling for the control mode in the $B^+ \to J/\Psi K^+$ fully selected data and simulation - with the selection placed on the constrained mass, and all weights applied. Run1 (top) and Run2 (bottom) are shown in three trigger categories: eTOS (left), hTOS (centre), and TIS (right).

4.13.3 HOP Modelling

The modelling of the m_{HOP} variable was studied using the control mode fully selected simulated samples and data. No significant biases are observed, as can be seen in Fig. 4.50 between simulation and fully selected data with the mass constraint selection applied.

4.13.4 $\cos\theta_K$ Study

In the rare mode the partially reconstructed background is dominated by decays of the type $B \to K\pi e^+e^-$. These decays do not only have contributions from $J = 1$ $K\pi$ systems (P-wave) such as the $K^*(892)$, but from $J = 0$ systems as well (S-wave). The $J = 1$ and $J = 0$ systems have different distributions in the $m_{K\pi}$ spectrum, and the helicity angle $\cos\theta_K$, where θ_K is defined as the opening angle between the kaon and the direction of the B computed in the rest frame of the K^* system. However, the simulated samples used to model the partially reconstructed background only include $J = 1$ decays of $B^{*+,0} \to K^*(892)e^+e^-$. The mixture of different wave state combinations is not known, so it is necessary to understand the maximal variation that different wave combinations could introduce to sensitive variables m_{Kee}, $\cos\theta_l$ and m_{HOP}. The first stage of investigation is

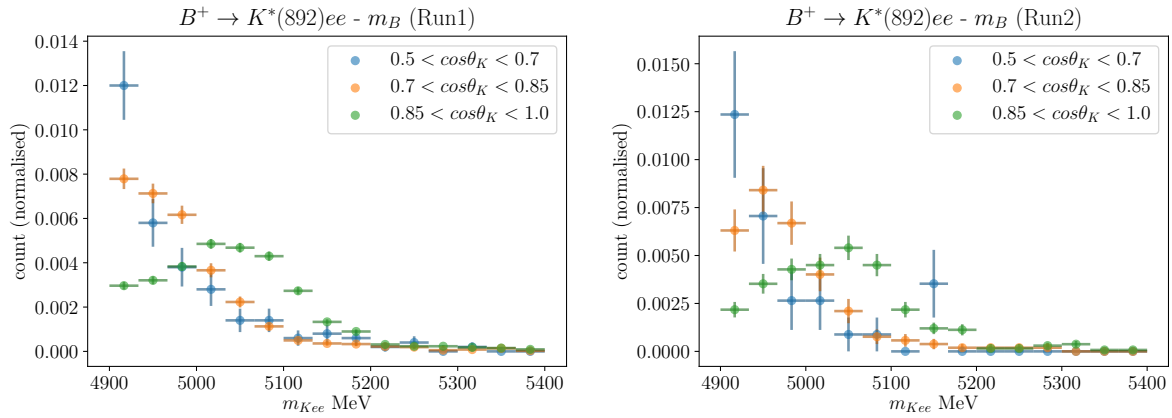

Figure 4.51: $B^+ \to K^*(892)e^+e^-$ simulated samples mass distribution in slices of true $\cos\theta_K$, for Run1 (left) and Run2 (right).

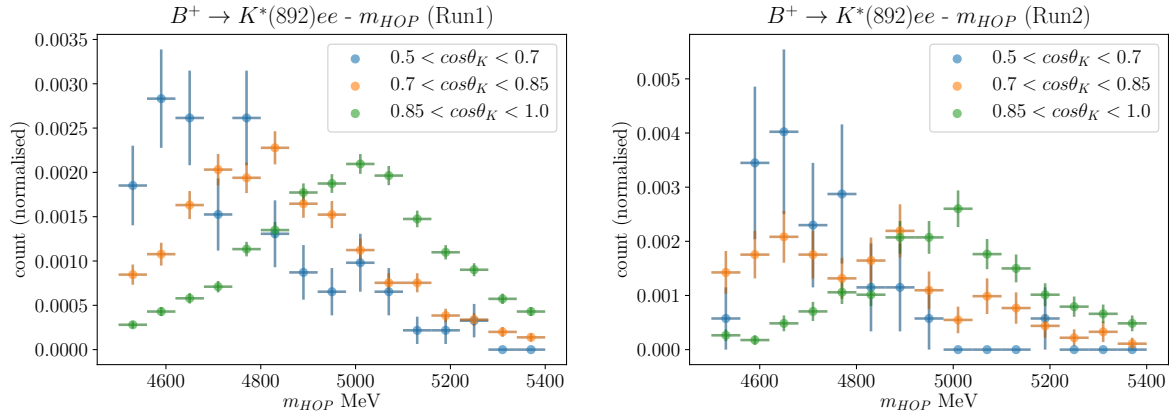

Figure 4.52: $B^+ \to K^*(892)e^+e^-$ simulated samples HOP mass distribution in slices of true $\cos\theta_K$, for Run1 (left) and Run2 (right).

performed using the helicity angle of the kaon $\cos\theta_K$. As each spin state has a different $\cos\theta_K$ distribution, the dependence of the sensitive variables of interest on $\cos\theta_K$ gives insight into how significantly they might change under a different mixture of spin states. Distributions are shown for m_{Kee}, $\cos\theta_l$ and m_{HOP} in Figs. [4.51, 4.52, 4.53] given in slices of $\cos\theta_K$. The m_{Kee} and m_{HOP} show clear variation as function of θ_K, whereas the angular distribution $\cos\theta_l$ does not.

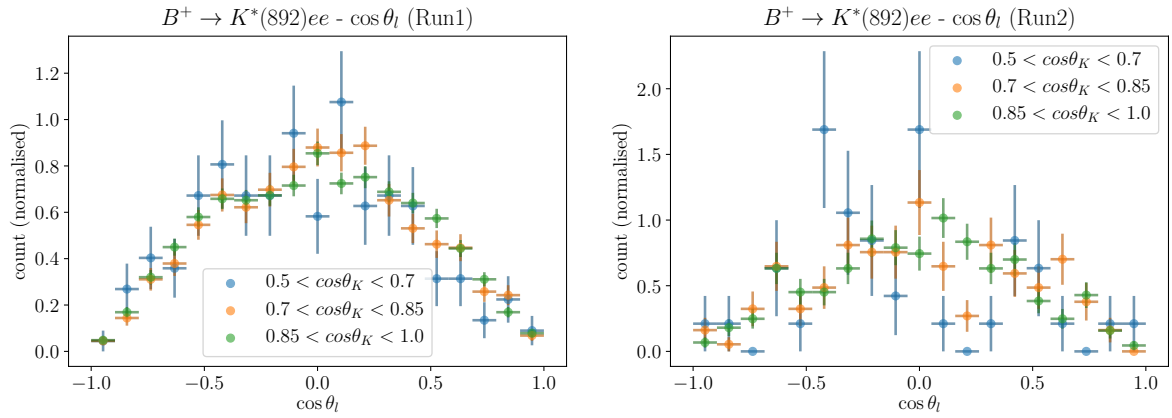

Figure 4.53: $B^+ \to K^*(892)e^+e^-$ simulated samples angular distribution in slices of true $\cos\theta_K$, for Run1 (left) and Run2 (right).

Two methods are compared to obtain the set of weights that would approximately transform fully spin state $J = 1$ simulated samples of $K^*(892)$ to $J = 0$, and use these weights to model the variation expected in sensitive distributions. The first involves taking the true $\cos\theta_K$ distribution from the $K^*(892)$ simulated samples with the full selection chain and acceptance performance applied and computing weights that flattened the distribution as would be expected at generator level for such decays. These weights are referred to as *'niave'*, as this method does not account for the effect of the acceptance that would naturally sculpt the $J = 0$ angular distribution in $\cos\theta_K$. Therefore a second method was investigated by deriving a weight that flattens the $\cos\theta_K$ prior to any acceptance or selection effects. By applying these weights to the $J = 1$, fully selected samples, $B \to K^*(892)e^+e^-$, the effect of the acceptance on a potential $J = 0$ contribution could be captured. These weights could then be sampled to convert a chosen percentage of the S-wave part-reco background to P-wave, and the impact on the mass and $\cos\theta_\ell$ distributions studied. The *'naïve'* weights and *'true'* weights are compared in Fig. 4.54 where the *'true'* weights can be seen to provide a much smaller correction, due to the weights correctly accounting for acceptance effects.

As the ratio of S/P wave composition is not well known, one must estimate a variation comparing 100% spin state $J = 0$, and 100% $J = 1$. This provides a maximal degree of variation from the nominal, which is found to be negligible, particularly in the context

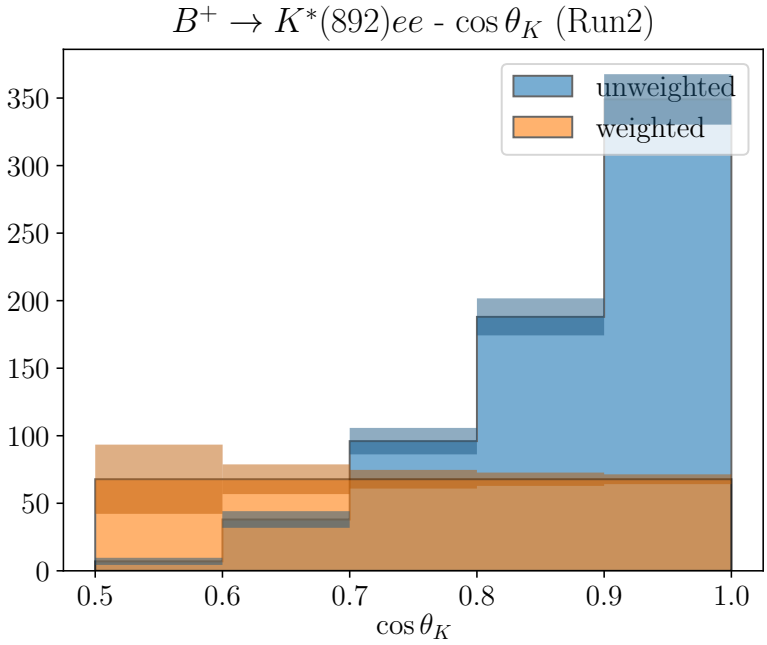

Figure 4.54: Run2 $B^+ \to K^*(892)e^+e^-$ simulated samples shown for the kaon helicity angle $\cos\theta_K$ (blue), weighted naively such that this helicity angle is flattened (green), and weighted accounting for true acceptance effects (orange).

of limited statistics in the rare mode fits. This study was performed in each trigger category separately, looking at both the mass and the HOP distributions. The difference between the naive and true weights is striking, and additional ratio plots are provided. The results of this second part of the study are found in Figs. [4.55, 4.56]. However, despite not knowing the ratio of background components, it is pertinent to understand the variation from S-wave to P-wave. Finally a set of kernel density estimate fits to the $K^*(892)$ simulated samples reconstructed mass distribution to investigate the impact of the spin state on the mass distribution used in the fit. These are shown as a scan from 0 - 100 % of the events weighted, and the fits are overlaid in Fig. 4.57. No significant variation is seen.

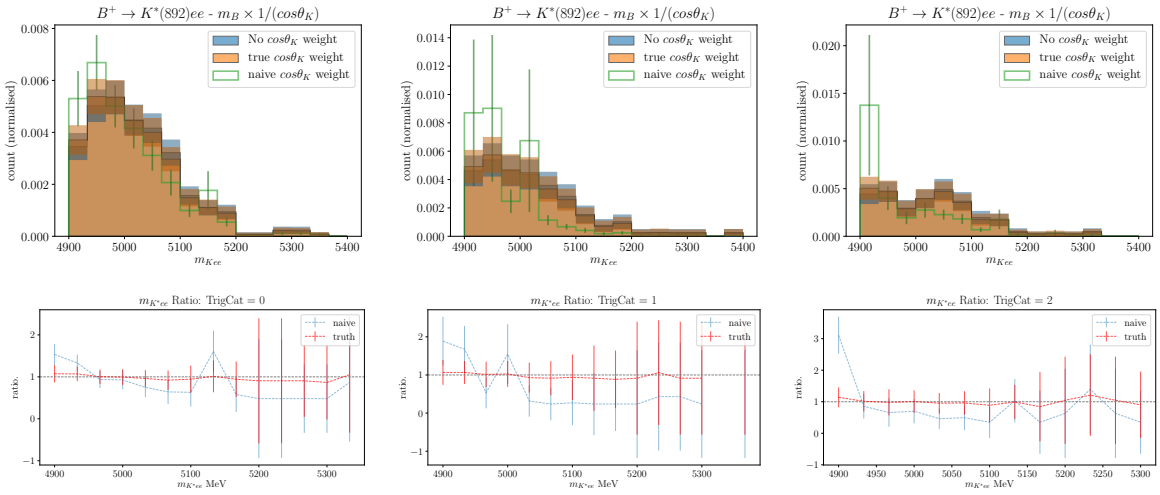

Figure 4.55: Run2 $B^+ \to K^*(892)e^+e^-$ simulated samples shown for the reconstructed mass m_B (blue), weighted naively such that the helicity angle is flattened (green), and weighted accounting for true acceptance effects (orange). Show for eTOS (left), hTOS centre), and TIS (right). Below each plot is a ratio plot of the weighted distribution against the unweighted distribution.

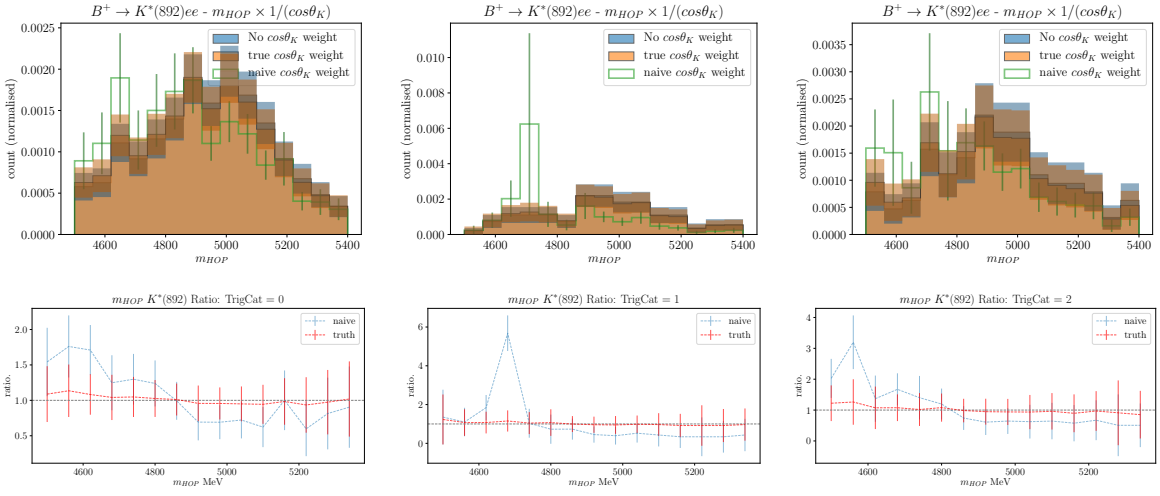

Figure 4.56: Run2 $B^+ \to K^*(892)e^+e^-$ simulated samples shown for the HOP mass m_{HOP} (blue), weighted naively such that the helicity angle is flattened (green), and weighted accounting for true acceptance effects (orange). Show for eTOS (left), hTOS centre), and TIS (right). Below each plot is a ratio plot of the weighted distribution against the unweighted distribution.

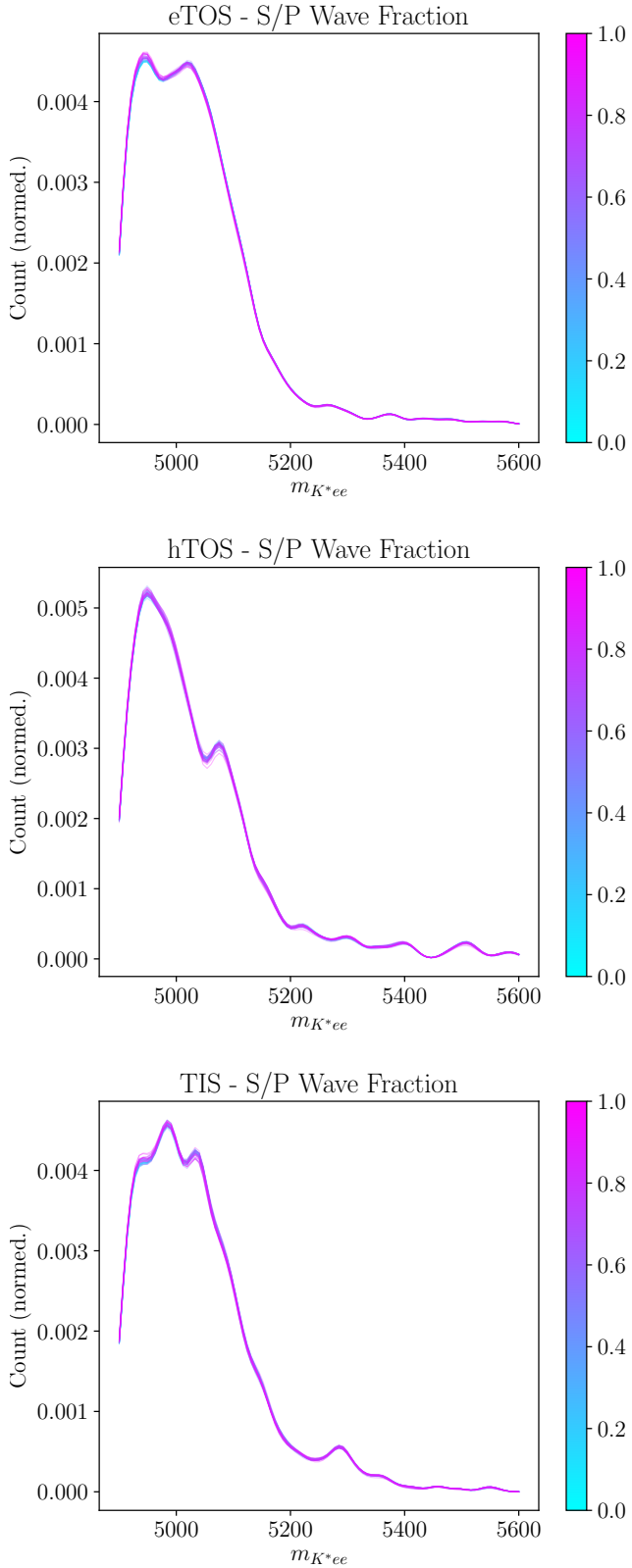

Figure 4.57: KDE fits to the $B^+ \to K^*(892)e^+e^-$ simulated samples with a fraction weighted as S-wave for increments from 0 - 100 %. This shows trends across the full mass spectrum and show that there is minimal variation for any ratio of S/P wave.

4.13.5 Full $\cos\theta_l$ for $K^+\pi^-$ system

The true $\cos\theta_l$ angular distribution for the $B \to K^*(892)e^+e^-$ decays accounts for the momentum of the Pion in the decay of the excited kaon, which by definition is not included in the angular distribution for $B^+ \to K^+e^+e^-$ decays. However, the difference in true angular distribution between the S and P wave decays could present a variation to the nominal angular distribution in this analysis. To investigate this the true 4-body angular distribution of $K^*(892)$ decays was obtained from `RapidSim` and a correction computed to model the S-wave contributions; this correction is used to weight the $K^*(892)$ simulated samples to understand the magnitude of not accounting for the fraction of events as S/P wave contributions. Additionally, the angular distribution accounting for the momentum of the pion was computed to show the difference when compared to the nominal 3-body method. This comparison can be seen in Fig. 4.58 where the true $K^*(892)$ angular distribution is denoted as `kst ctl`. The impact of introducing increasing fractions of S-wave into the model is shown in Fig. 4.59 where four fractions are shown from 0% to 100% S-wave. For fractions less the 50% the effects are minimal. The impact of these weights on the other key distributions; the m_{Kee} and the m_{HOP}, are shown in Fig. 4.60 and Fig. 4.61 where a comparison between no weights, the $\cos\theta_K$ weights used in Sec. 4.13.4, and the weights obtained from the true $K^*(892)$ $\cos\theta_l$ distribution. These plots are all shown for maximal S-wave mixing of 100% and the effects are negligible.

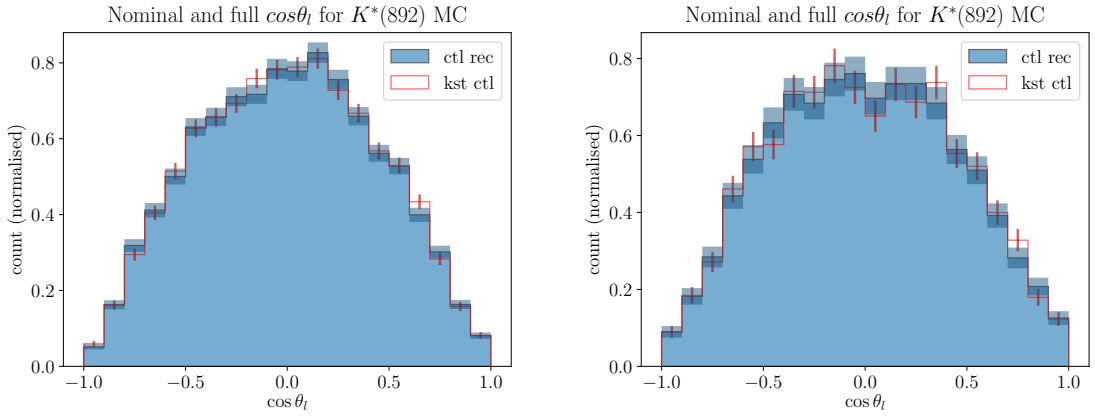

Figure 4.58: Comparison of nominal 3-body $\cos\theta_\ell$ (ctl rec) and the true 4-body angular distribution (kst ctl) computed for the $B \to K^*(892)e^+e^-$ simulated samples. Run1 (left) and Run2 (right).

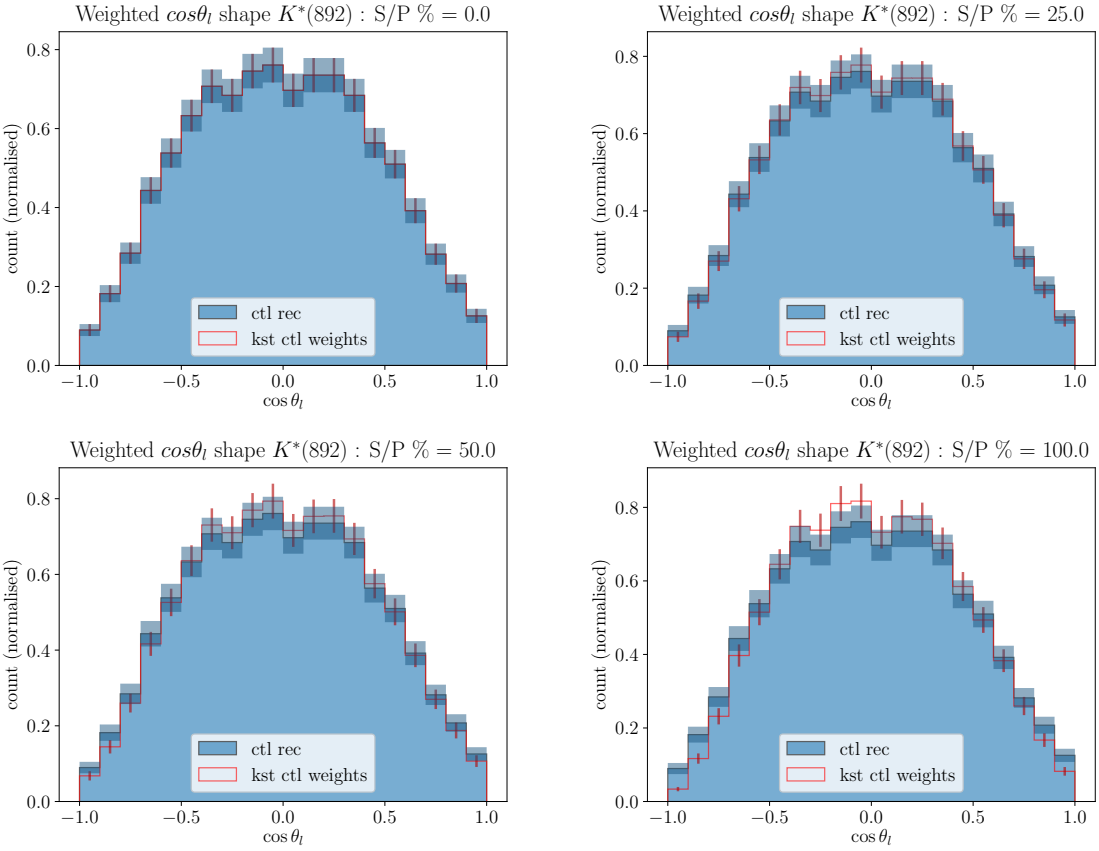

Figure 4.59: Modelling the S-wave contribution to changing the angular distribution of $B \to K^*(892)e^+e^-$ simulated samples. Shown here in red the S wave composition of 0%, 25%, 50% and 100% against the nominal fully P-wave shape in blue. The sculpting effect of even 50% S-wave is small.

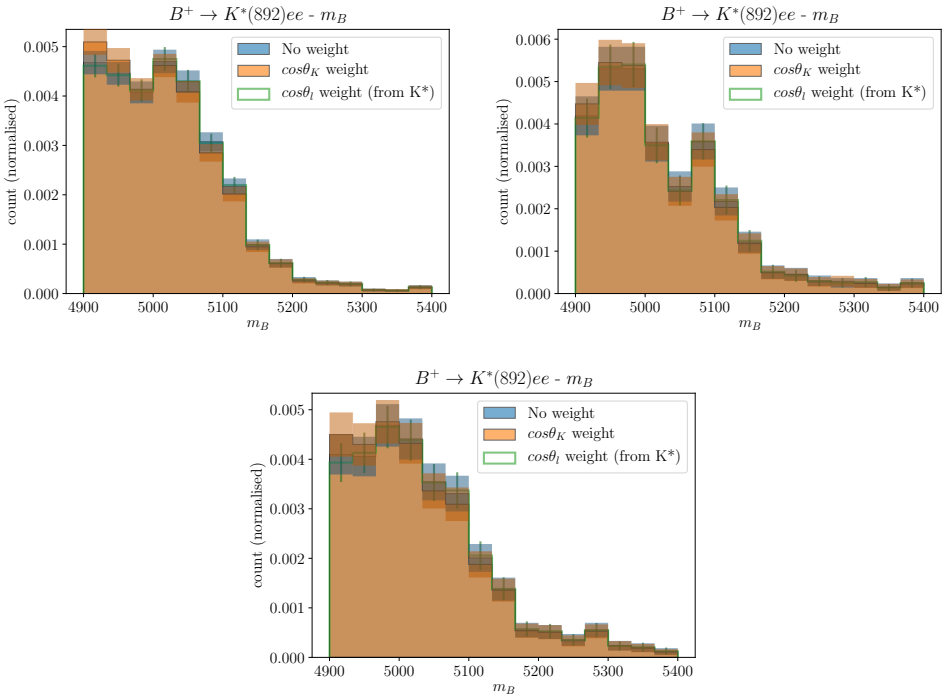

Figure 4.60: Mass distribution for $B \to K^*(892)e^+e^-$ simulated samples, comparing the nominal distribution with no weights in blue, with the $\cos\theta_K$ weights in orange, and with the full 4-body $\cos\theta_l$ weights from the $K^*(892)$ in green. Shown for eTOS(top left), hTOS (top right) and TIS (bottom).

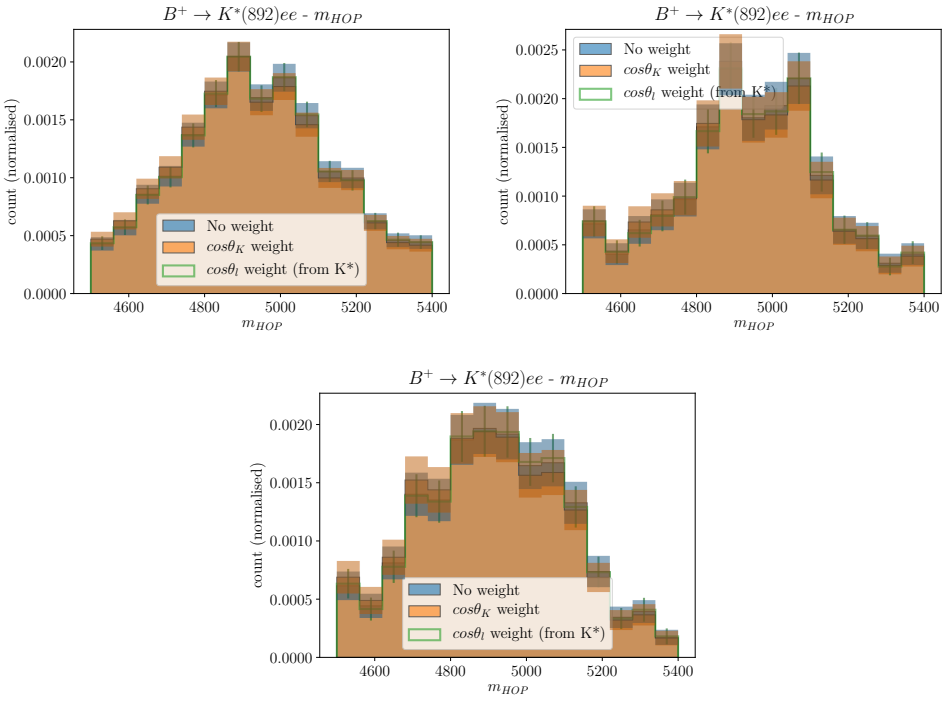

Figure 4.61: HOP mass distribution for $B \to K^*(892)e^+e^-$ simulated samples, comparing the nominal distribution with no weights in blue, with the $\cos\theta_K$ weights in orange, and with the full 4-body $\cos\theta_l$ weights from the $K^*(892)$ in green. Shown for eTOS(top left), hTOS (top right) and TIS (bottom).

	eTOS	hTOS	TIS
Run1	0%	0%	0%
Run2	0.4%	0.5%	0%

Table 4.15: Details of the presence of multiple candidates in the $B^+ \to K^+e^+e^-$ dataset for Run1 and Run2.

4.13.6 Multiple Candidates

The multiple candidates are found in the data when the `eventNumber` and `runNumber` branches do not provide a unique combination. These multiple candidates present a negligible contribution to the fully selected $B^+ \to K^+e^+e^-$ dataset, being present maximally at the sub percent level in all trigger categories, in both Run1 and Run2. Full details of the study are given in Table.4.15.

Chapter 5

RICH Mirrors

In this chapter the alignment of the RICH mirror detectors will be presented, along with the authors contribution to the real time alignment framework and monitoring of the RICH mirrors. Additionally a brief description of the mirror testing procedure is shown and the update to measurement algorithms as produced by the author.

At the LHCb experiment the two RICH detectors provide vital information used to perform particle identification, details of which are discussed in Chapter. 3. The RICH mirror system is made up of a total of 100 mirrors, each of which must be correctly aligned. This requires two pieces of information, firstly a measure of the physical position of the mirror relative to the rest of the detector, and seconadly an understanding of the surface roughness.

5.1 Alignment Method of RICH1 and RICH2 Mirror Systems

The ring imaging Cherenkov detectors provide particle identification information on track candidates that is used in all physics analyses at LHCb. As outlined in Sec 3.3.4.1 this

relies on the unique distribution of Cherenkov ring opening angle as a function of particle momentum that allows separation between various particle candidates. The cosine of the Cherenkov angle is:

$$\cos\theta_C = \frac{1}{n\beta} = \frac{1}{n}\frac{\sqrt{m^2+p^2}}{p}, \tag{5.1}$$

where n is the refractive index, m the mass of the particle, p the momentum of the particle. Considering Fig. 3.13 in the light of Eq. 5.1 it is apparent that at least two detectors with different refractive mediums are needed, as at higher momentum the opening angles converge and the system becomes saturated. The momentum and at angle at which saturation occurs is a function of the refractive index of the material.

Events have high occupancy and therefore many particles passing through the RICH detectors, this multiplicity results in a large degree of overlap between Cherenkov rings. To separate the rings a log-likelihood minimisation is used. The first stage of the algorithm uses ray-tracing to associate photon hits with a charged track. The photon is assumed to come from the middle of the charged track as it is not possible to assert otherwise, and a systematic uncertainty σ_θ assigned as the resolution. The second stage calculates the log-likelihood under different mass hypothesis for each track in the event. As this is computationally expensive, tracks removed from the iteration pool once a clear preference for particle type is found.

Knowing the exact physical locations and orientations of the mirrors is vitally important as the ray tracing to associate photons with tracks relies on the optical geometry of the mirror system. Additionally a mis-aligned system increases the resolution σ_θ, degrading the quality of the probability predictions. A worse resolution also impacts computational load as more tracks have to be considered for each track.

5.1.1 Misalignment Identification

To perform the alignment of the RICH mirror system any physical deviations of the RICH mirrors from their locations as specified in the LHCb software must be accounted for. In a perfectly aligned system the Cherenkov photons form a perfect circle with the projection

of the charged track at the centre of the ring. If one, or a number, of the mirrors have a small deviation the ring will be shifted with respect to the central projection of the charged track. The effect of a misalignment is best understood pictorially, as shown in Fig. 5.1. A shift in the primary mirror of some small angle α results in a shift of 2α on the detector plane, resulting in a Cherenkov ring shifted with respect to the projected position of the track at the centre. The shift, δ, is the difference between the measured Cherenkov angle θ and the expected angle θ_C

$$\delta\theta(\phi) = \theta(\phi) - \theta_C \tag{5.2}$$

as a function of the azumithal angle ϕ.

When the detector is well aligned the Cherenkov angle is independent of the azimuthal angle, however, when misaligned there is a sinusoidal dependence. Therefore:

$$\delta\theta(\phi) = \Theta^y \sin(\phi) + \Theta^z \cos(\phi), \tag{5.3}$$

where Θ^y and Θ^z are the misalignment on the detector plane. An example is shown in Fig. 5.2 where the two dimensional histograms show $\delta\theta$ against ϕ, where the expression in Eq. 5.3 is fitted showing the clear sinusoidal and flat shapes corresponding to misaligned and aligned systems. The distributions of $\delta\theta$ for RICH1 and RICH2 are shown in Fig. 5.3.

5.1.2 RICH Mirror Alignment Procedure

The alignment procedure is a data-driven method, and while a flowchart of the steps is shown in Fig. 5.4 for reference, the important steps are detailed below.

1. The latest alignment conditions are taken from the LHCb database, and a sample of pre-selected events are obtained.

2. The pre-selected tracks are reconstructed and filtered to have a momentum of greater than 20 GeV for RICH1, and greater than 40 GeV for RICH2. This ensures saturation of Cherenkov angle.

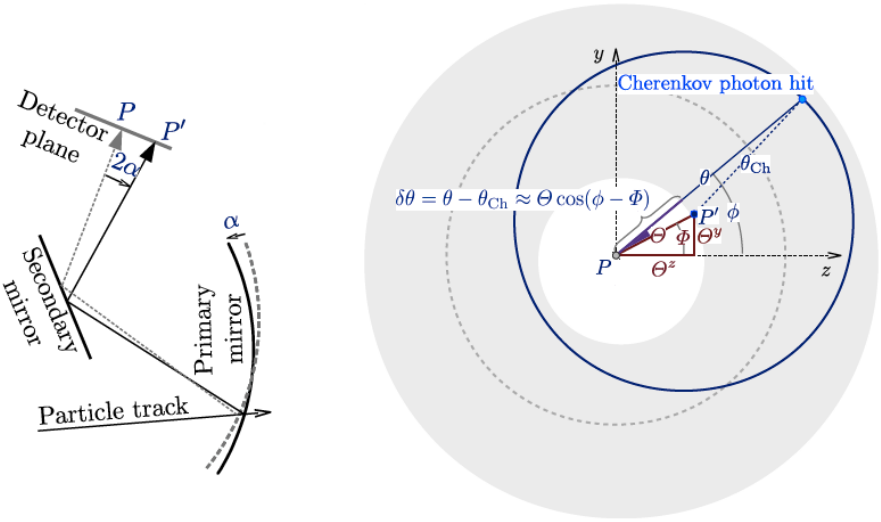

Figure 5.1: Variation in the primary mirror alignment by some angle α results in a 2α variation on the detector plane. (left) The Cherenkov ring shifts with respect to the particle hit location due to a misalignment. (right)

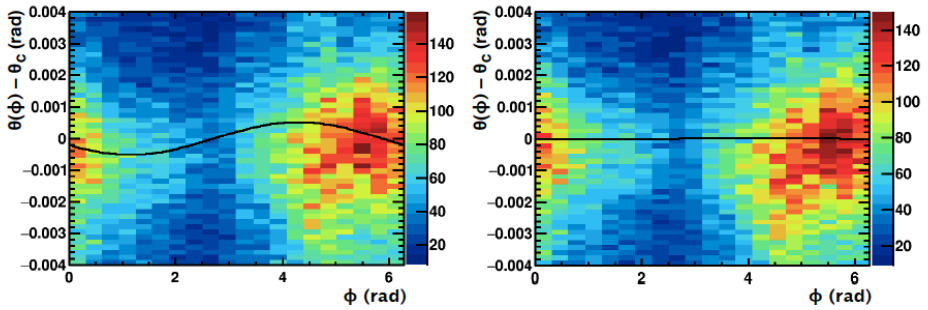

Figure 5.2: A misaligned system results in a sinusoidal distribution in ϕ (left) while a fully aligned distribution is flat (right). The black line is the resulting fit to the $\delta\theta_C$ histograms, examples of which are given in Fig. 5.3.

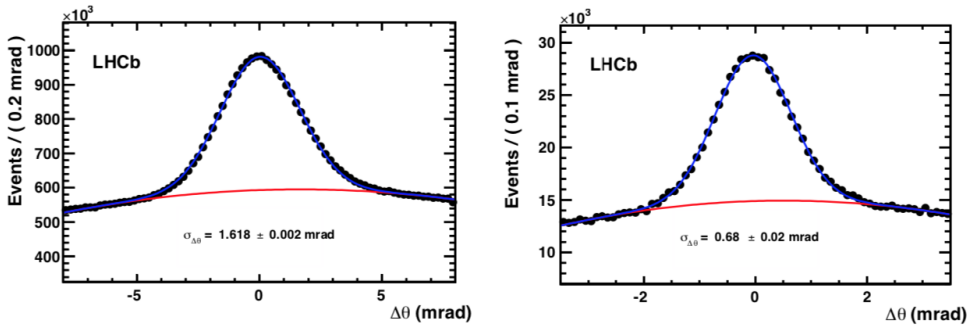

Figure 5.3: RICH1 and RICH2 $\delta\theta_C$ distributions. [30]

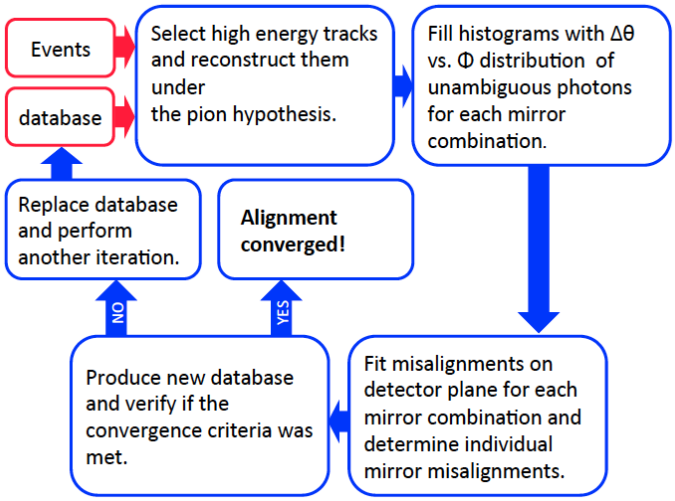

Figure 5.4: Flowchart of alignment procedure, which iterates until the convergence criteria is reached.

3. Photons reaching the HPDs could have come from any combination of primary and secondary mirror combination, so histograms of $\delta\theta$ vs. ϕ are filled for each. A sinusoidal function is fitted to the histogram, an example of which is shown in Fig. 5.2.

4. The individual misalignment for each mirror is obtained through the solution of a system of linear equations.

5. If the convergence criterion is met, the LHCb alignment database is updated with the new mirror orientations. If not another iteration is performed using the alignment parameters found in the previous iteration. This is repeated until convergence is reached.

To fill the $\delta\theta$ vs. ϕ histograms, each track that passes the selection is projected onto the detector plane, and the hits in the detector around the projected location are taken as Cherenkov photon candidates. As there is no method to determine the emission point of the photons, it is assumed they were emitted in the middle of the track detector volume. This data is cleaned by selecting photons that, regardless of the exact emission point, will always reflect from the same mirror combination pair. The expected θ_C is calculated for the pion hypothesis and $\delta\theta$ is computed.

The misalignments Θ^y and Θ^z are caused by rotations of both the primary and secondary mirrors when compared to their positions in the LHCb software. Using the small angle approximation the misalignments can be expressed in terms of rotations α_p^y, α_p^z of the primary mirrors, and rotations β_s^y, β_s^z of the secondary mirrors, around y and z as:

$$\Theta^y = A_{ps}^y \alpha_p^y + B_{ps}^y \beta_s^y + a_{ps}^z \alpha_p^z + b_{ps}^z \beta_s^z, \tag{5.4}$$

$$\Theta^z = A_{ps}^z \alpha_p^z + B_{ps}^z \beta_s^z + a_{ps}^y \alpha_p^y + b_{ps}^y \beta_s^y. \tag{5.5}$$

The major magnification coefficients A_{ps}^y, B_{ps}^y, A_{ps}^z, B_{ps}^z and minor magnification coefficients a_{ps}^y, b_{ps}^y, a_{ps}^z, b_{ps}^z translate the effects of a mirror rotation onto the detector plane. These magnification coefficients are calculated empirically during a technical stop as the process takes $\mathcal{O}(10)$ times longer than the normal alignment. The combination of primary and secondary mirrors results in an under constrained system where there are 32 equations (for RICH1) and 40 unknowns. An absolute misalignment could be the result of either a primary mirror or secondary mirror deviation, or a contribution from both. The system of equations is solved using an L2 regularisation method, which is a least squares fit with an additional term included in the cost function to be minimised. The L2 method always returns a single stable solution, and treats every mirror with the same weight. For RICH2 there are approximately 1000 physically possible combinations of primary and secondary mirrors, however it is possible to choose a set of 96 such that all mirrors are included and that the $\delta\theta$ vs. ϕ histograms are suitably populated.

In 2015 and 2016 the alignment of the RICH detectors was manually operated. Data was collected and once the required amount was obtained (1M events for RICH1 and 2M events for RICH2) the alignment was automatically run, however any update of the database was done by an alignment expert on shift either at the LHCb pit or remotely. The resolution of Cherenkov rings in RICH1 is 0.88 mrad, and 0.5 mrad in RICH2, therefore an alignment tolerance of significantly less than this is desired to promote stability[51]. In Fig. 5.5 and Fig. 5.6 the alignment for all individual mirrors for RICH1 and RICH2 can be seen. The dashed lines show the conservative thresholds during 2016, which contain

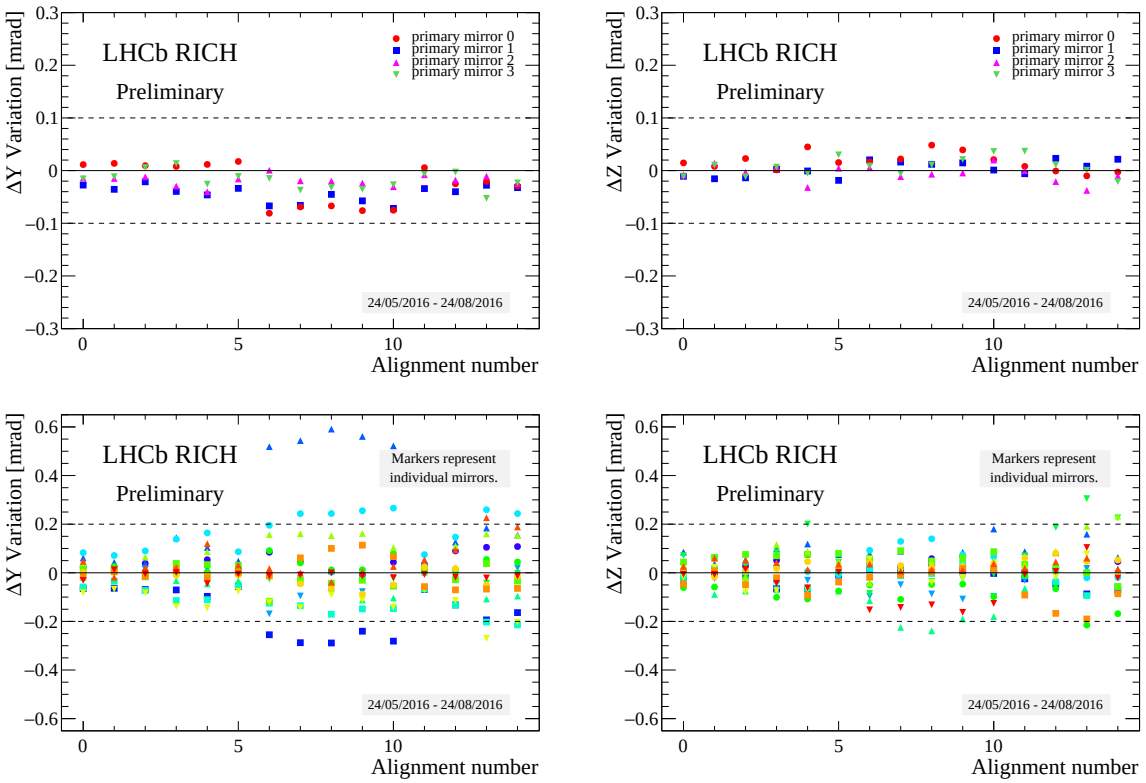

Figure 5.5: Variation in ΔY (left) and ΔZ (right) for Primary (top) and Secondary (bottom) RICH1 mirror alignment. (Performed on runs taken between 24/05/2016 and 24/08/2016). Each point represents an individual mirror rotation.

the majority of mirror rotations. Mirror rotations with respect to the database values in the y and z planes are given as Δy and Δz respectively. The Δy plots for RICH1 both primary and secondary mirrors show a consistent variation between run 5 and 10, however no update was required, and RICH2 maintained stability.

5.2 Real Time Alignment of RICH1 and RICH2

In 2017 the RICH alignment was moved fully online; this automated the update of alignment constants in the database such that it could be run in real time. The fully online trigger system at LHCb means that the alignments must be correct at time of data collection, as the data cannot be re-processed. An automatic updating procedure must balance the needs to have a fairly stable database that is not updated too frequently as this could obscure underlying faults in the alignment, with the need for regular updates

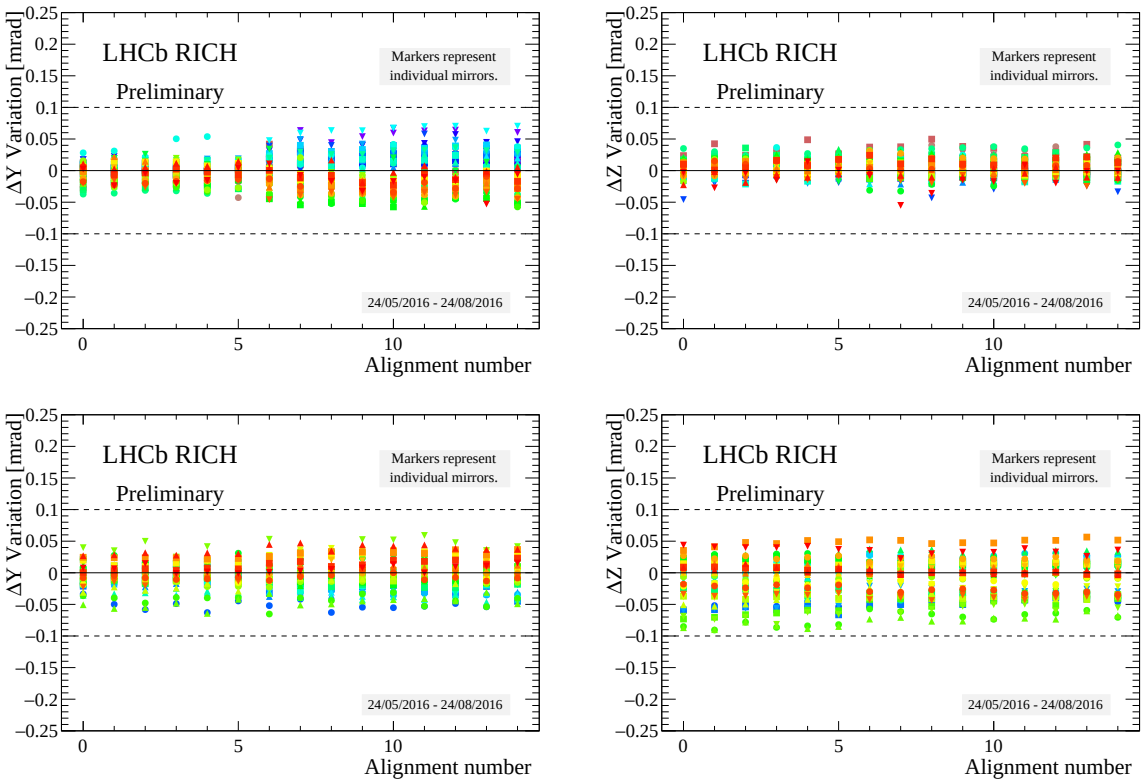

Figure 5.6: Variation in ΔY (left) and ΔZ (right) for Primary (top) and Secondary (bottom) RICH2 mirror alignment. (Performed on runs taken between 24/05/2016 and 24/08/2016). Each point represents an individual mirror rotation.

to maintain the alignment to a high quality.

To move this procedure online required the following steps:

- The variation in alignment constants over a typical run period needed to be assessed;
- A threshold chosen that ensured the highest quality alignment;
- A threshold chosen that preserved database stability.

RICH1 and RICH2 are updated independently but follow identical procedures. In each RICH detector the primary and secondary mirrors are updated synchronously, and the alignment constants in the database are updated when the relative shift in any mirror of the primary and secondary set is greater than the threshold.

There is no obvious experimental reason that should cause a shift to the calibration constants, and the physical mirrors themselves should not move, however in Run1 a notable step shift was seen in the calibration constants between the magnet up and magnet down

RICH	Mirrors	$\lvert\Delta Y\rvert$	$\lvert\Delta Z\rvert$
RICH1	Primary	0.030	0.030
RICH1	Secondary	0.046	0.037
RICH2	Primary	0.030	0.030
RICH2	Secondary	0.050	0.060

Table 5.1: Thresholds for RICH1 and RICH2 mirror misalignments.

running conditions. This was initially thought to be caused by potential leakage of the magnetic field into the RICH1 enclosure. This effect was not seen in Run2 where the only difference was an update to the magnetic field corrections, however, despite this it was decided to update the alignment constants each time the magnetic field updated. This has two effects, firstly to ensure that any unforeseen changes between magnetic field polarities would be accounted for, and secondly that a regular update to the calibration conditions was included for safety.

Using historical data from 2015/2016, thresholds were calculated such that the mirror alignment results in no update for 95% of fills. This was chosen to balance regular updates and stable running conditions. The table in Table. 5.1 shows the alignment threshold values for $|\Delta Y|$ and $|\Delta Z|$ that were used during 2017 and 2018 data collection periods. These values are comfortably within the required limits for PID performance.

The RICH alignment software was updated to include these tolerances and to automatically update the alignment constants in the LHCb database. In addition, the updated alignment plots were sent directly to the control room to provide real time feedback to shift workers. In Figs.[5.7, 5.8] data from 2017 is shown for both rich detectors, and in Figs.[5.9, 5.10] data from 2018 is shown for both rich detectors.

The authors contribution to the software was to update the alignment package and incorporate it into the online framework such that it could be called automatically in the LHCb real time stack. The update of the alignment values in the database is represented in Figs.[5.7 - 5.10] as the points where the alignment marks collapse together onto the y axis showing that all mirrors are in an identical orientation to the database values.

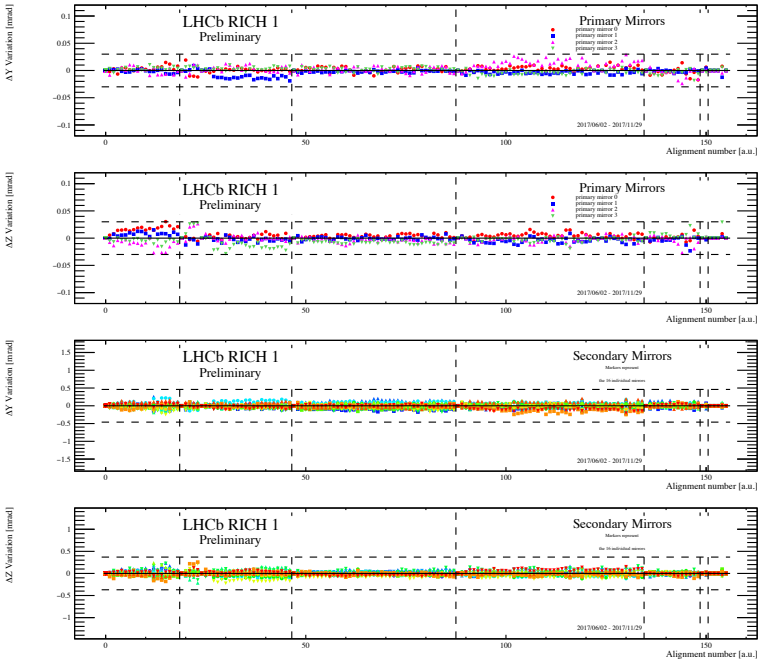

Figure 5.7: Variation in ΔY and ΔZ for primary (top pair) and secondary (bottom pair) mirror alignments for RICH 1 during 2017 data taking. Vertical dashed lines represent magnet polarity switches, horizontal dashed lines represent thresholds. Each point represents an individual mirror rotation.

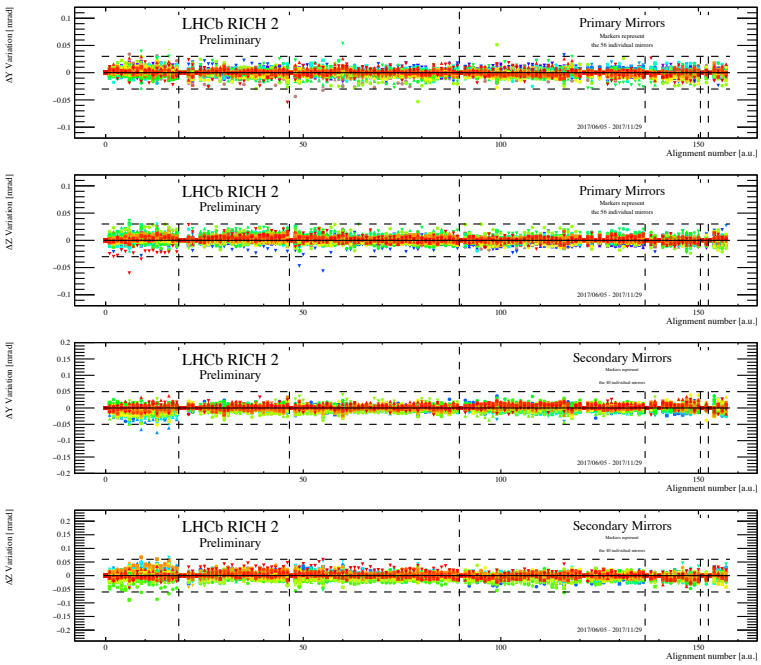

Figure 5.8: Variation in ΔY and ΔZ for primary (top pair) and secondary (bottom pair) mirror alignments for RICH 2 during 2017 data taking. Vertical dashed lines represent magnet polarity switches, horizontal dashed lines represent thresholds. Each point represents an individual mirror rotation.

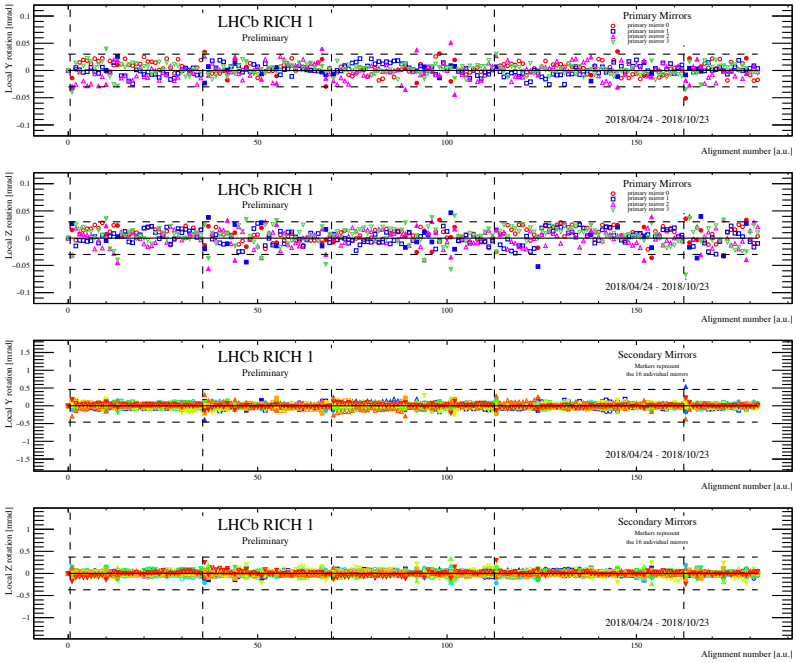

Figure 5.9: Variation in ΔY and ΔZ for primary (top pair) and secondary (bottom pair) mirror alignments for RICH 1 during 2018 data taking. Vertical dashed lines represent magnet polarity switches, horizontal dashed lines represent thresholds. Each point represents an individual mirror rotation.

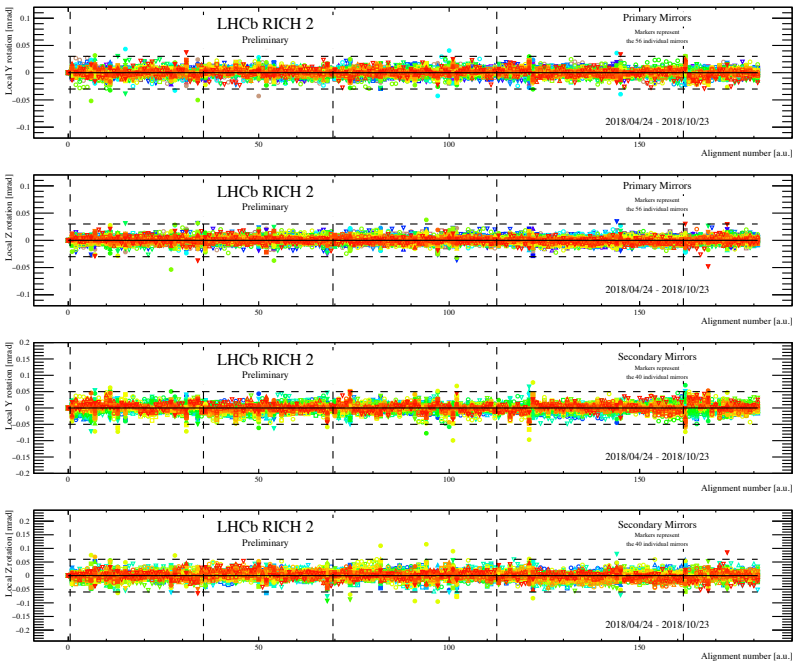

Figure 5.10: Variation in ΔY and ΔZ for primary (top pair) and secondary (bottom pair) mirror alignments for RICH 2 during 2018 data taking. Vertical dashed lines represent magnet polarity switches, horizontal dashed lines represent thresholds. Each point represents an individual mirror rotation.

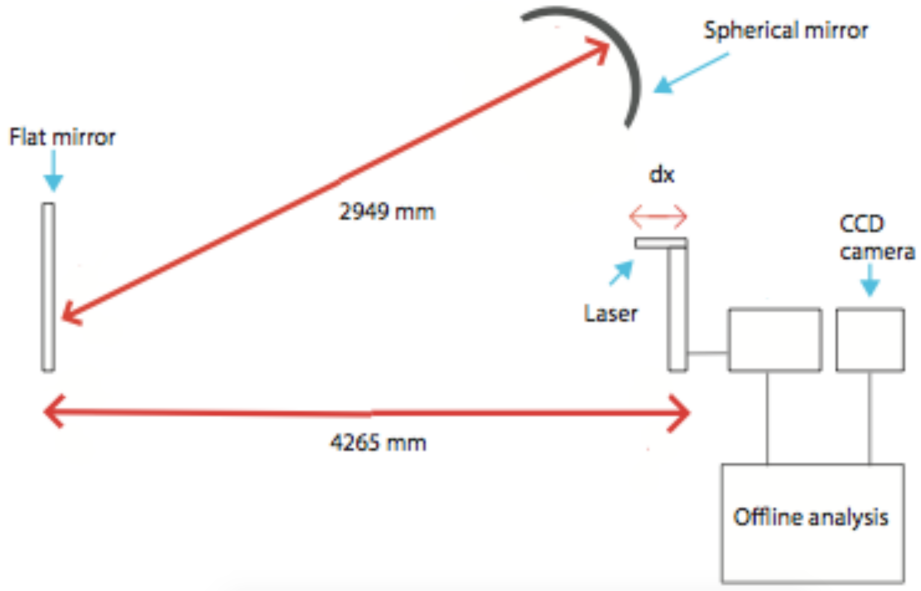

Figure 5.11: Diagram of the optical table set up used to measure the $D0$ of the prototype mirrors. The laser is reflected off the flat mirror (either the reference or test mirror) to a curved mirror to focus the beam back onto the CCD of the digital camera via a second reflection off the flat mirror. The camera was mounted such that it could be smoothly moved along a track with mm precision.

5.3 RICH Mirror Prototype Testing

The design specification of mirrors used in the RICH detectors are an unusual technical challenge, and custom prototypes were tested in the dark room lab at CERN to ensure quality control. This section will focus on one specific aspect of the prototype testing, the updating of the measurement of the $D0$ and preliminary testing of the flat glass mirrors to quantify the surface quality.

By using the sample mirror to focus a laser and measuring the size of the laser spot the surface quality can be characterised in a comparable way. The size of the focused laser spot is defined as the $D0$, a width of a circle that contains at least 95% of the focused light. This is typically done by drawing circles of increasing radius around the centre of the focused spot until 95% of the integrated intensity of the image is contained. This is by definition a conservative estimate of the true $D0$, but for a strong laser or a low background image is sufficient. This method relies on three assumptions:

1. **That the focused laser spot is a circle** - if the laser is no longer focused as a

circle there is some defect in the mirror and this results in a larger $D0$ measurement;

2. **That the centre of the focused spot is used to draw enclosing circles** - this requires a suitable peak finding algorithm. This was designed not simply to take the highest pixel value, but to find the centre of the most prominent peak;

3. **That the background level in the image is negligible** - if there is a non-negligible background, the number of background pixels dramatically outweigh the total integrated intensity from the laser due to saturation of the CCD. Therefore the background needs to be modelled and subtracted in a consistent and automatic way.

The dark room set up to measure the $D0$ is shown in Fig. 5.11. The laser is reflected off the testing mirror onto a reference spherical mirror that serves to focus the beam, which is reflected a second time off the testing mirror and onto the CCD of a digital camera placed next to the laser source, that is attached to a rail so that it can be smoothly moved. The testing mirror is moved to roughly focus the laser onto the CCD, then during data taking the camera is moved along the rail by fixed increments between each exposure to find the focus point of the system to the closest mm. The entire system is housed in a dark underground lab and any residual light from computer monitors was shielded by a series of heavy curtains.

Due to the camera lacking any cooling system the background required subtracting. The average level of the background was found to fluctuate significantly even between successive dark images taken with no laser source present, which meant subtracting a background template was not viable. It was found that within several cm of the laser focus the peak was always significantly greater than the background, and as this was sufficient to cover the distance required to find the focus a lower threshold was placed such that the background was always removed. Studies showed that the absolute level of this background did not alter the $D0$ value until an order of magnitude higher than the observed background level. This can be seen in the data presented in Fig. 5.12. A

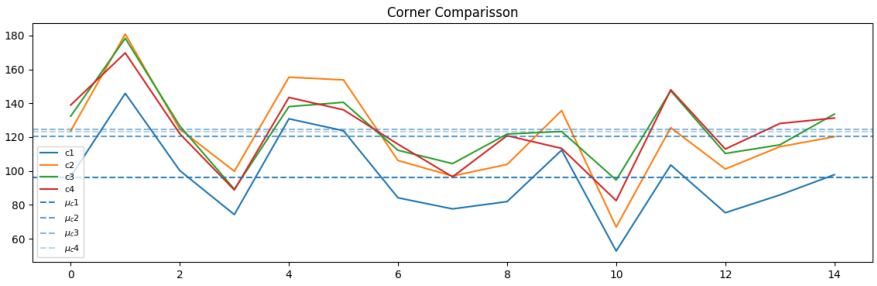

Figure 5.12: Average pixel value for four quadrants of an image taken with the laser off. The quadrants move synchronously but with a large degree of variation. This data shows 15 images taken over the course of a day.

conservative threshold was chosen and all pixel values with intensity less than the threshold were clipped to 0.

The performance of the algorithm used to find the $D0$ size required an updated along with the new equipment. Drawing concentric circles on a pixel grid without double counting any pixels is not a trivial task, and the method was significantly developed to increase the performance. The method exploited a combination of bit-wise shifts and caching to efficiently increase the size of the $D0$ circle without double counting or fully recomputed the $D0$.

Reference Mirror

In the lab there is a well calibrated flat reference mirror used to obtain a baseline performance of the method. In Fig. 5.13 two example $D0$ measurements are shown, on the left a poorly focused result, and on the right a well focused result, where the well focused image shows a clear reduction in the width of the circle. The relative prominence of background in the poorly focused image is due to the relative lower intensity of the peak to background. The results of scanning around the focal point is shown in Fig. 5.14 where a clear minima is found. This minima gives the $D0$ as approximately 0.0001 m. Further work was subsequently conducted to implement and update methods on the real samples once they arrived from the manufacturers and should be installed in the LHCb detector during 2020.

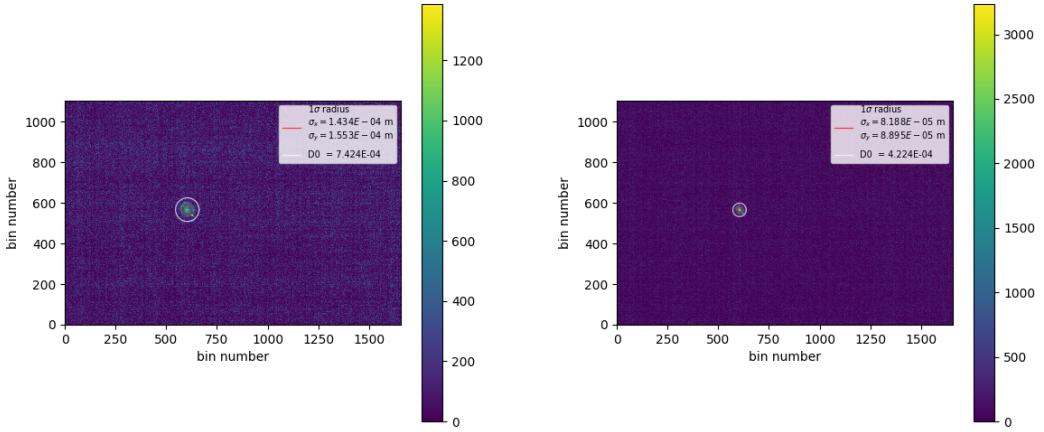

Figure 5.13: Reference mirror example D0 measurements at two example stages of focus.

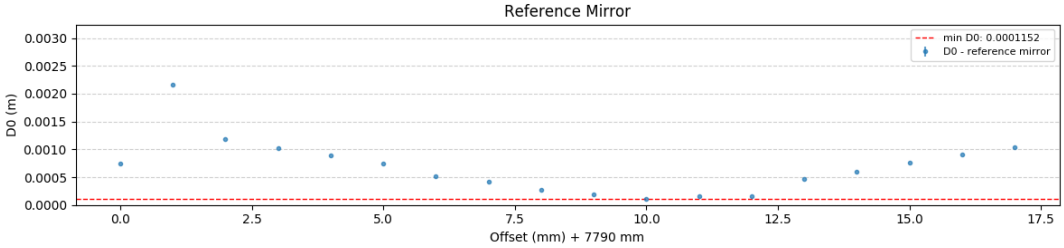

Figure 5.14: D0 results from scan around the focus point in 1 mm increments.

Sample Olomouc Mirror

The company Olomouc were tendered to produce the flat glass mirrors, and in 2018 a prototype sample was brought to CERN to be measured for quality control. The experimental procedure was identical to that using the reference mirror where the focus point was found and the D0 measured. As is readily apparent in Fig. 5.15 the pattern observed does not resemble the clear single spot seen on the reference mirror in Fig. 5.13. Rather than being mono-modal there is a clear bi-modal distribution that is indicative of poor surface smoothness. This kind of optical effect on the scale observed was insufficient for the mirrors for RICH1. However, in order to formally deal with the contractor a full measurement needed to be made to demonstrate the D0 was significantly greater than the contracted specifications.

From studies observing the variation in background level with the laser off, as shown

in Fig. 5.12, an estimate of the maximum background level could be obtained. Images were taken with the laser off, and the dark room in operating condition, over the course of a day and each image segmented into four quadrants. The average pixel values for each quadrant are plotted. What is clear is that the variation is substantial making a single background template that would be subtracted from laser-on images invalid. Therefore, the background was removed by clipping all pixel values less than 200. This has a small impact on the absolute value of the D0, but is a much smaller effect than the inflation of the D0 measurement by the high background level. This method was validated on the sample mirror, achieving a consistent $D0$ measurement.

To demonstrate that the magnitude of the D0 was a result of the mirror rather than the analysis method a scan of background subtraction was performed. This involved starting with no background subtraction and scanning through increasing percentages of the expected background until a background subtraction of 10× greater than expected was being performed. These results are shown in Fig. 5.16 where the D0 size can be seen to decrease until it plateaus above 200% of the expected background. This demonstrates that regardless of the background subtraction the D0 size of the prototype mirror did not meet specification, being measured at least to be greater than 11 mm, while the design specification for the RICH1 flat mirrors stipulated 2.5 mm[51].

This project is part of the ongoing contribution of the LHCb group at Bristol to the running of the LHCb experiment. Work on the RICH mirrors has been shared between PhD students, and the upgrade of the mirrors has been passed on to the next round of students. Following these studies the company has provided 20 mirrors that meet or exceed specifications. Both the spherical composite mirrors and these flat mirrors have passed the quality assurance tests. The glass mirrors have been coated with a reflective coating and are currently being prepared for installation into the RICH1 sub-detector. While the spherical mirrors are waiting to be coated by the CERN thin-film and glass coating group.

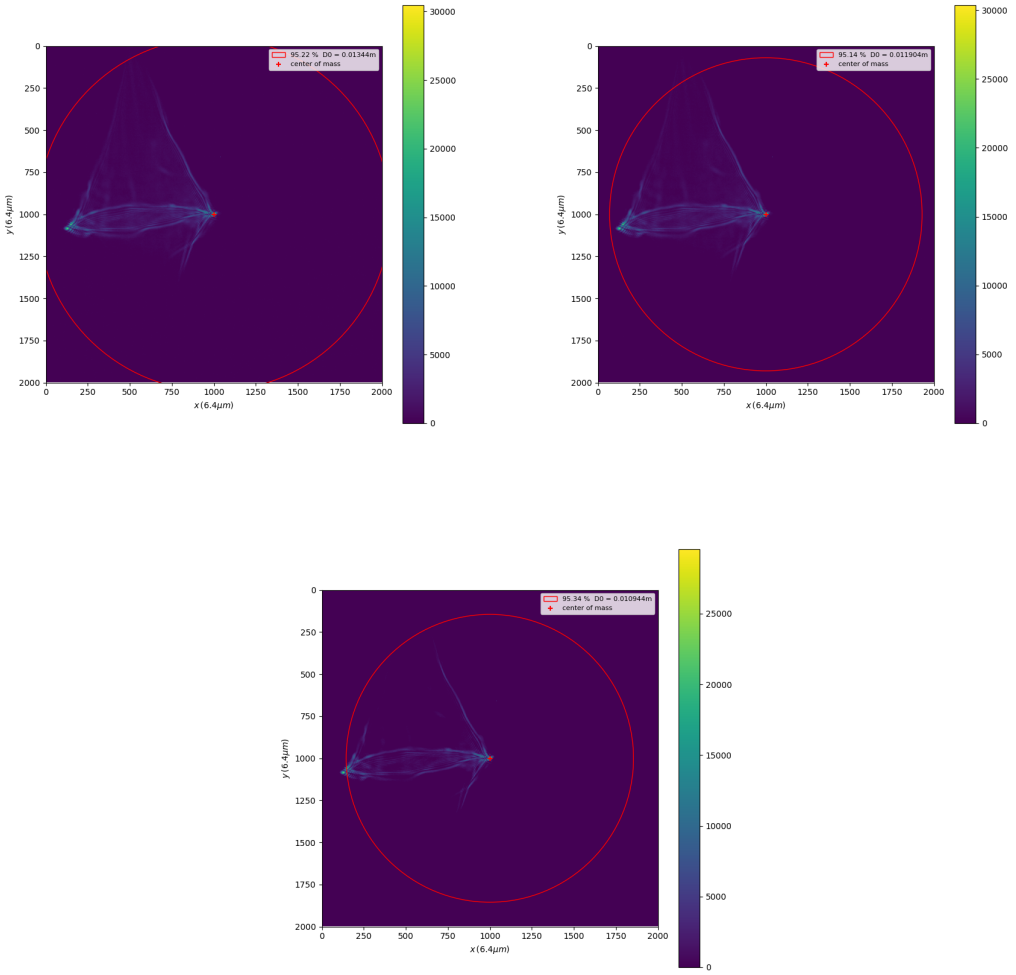

Figure 5.15: Background subtracted from prototype image at best focus point. 0x expected background removed (top left), 1x expected background removed (top right), 9x expected background removed (bottom). Bimodal structure remains clear in all cases. Degradation clear compared to the reference mirror shown in Fig. 5.13.

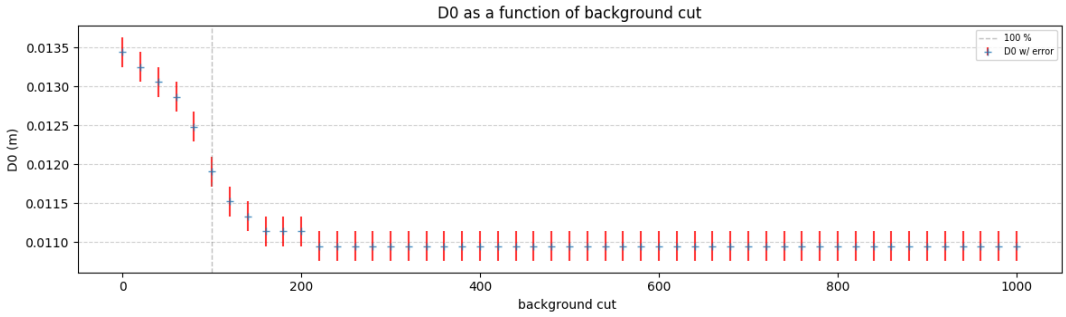

Figure 5.16: Scan of fitted D0 size over a range of 0× expected background removed to 10× expected background removed. Plateau shows the result is not caused by the residual background distorting the result. The minimum D0 of this mirror was 11 mm.

Chapter 6

Track Reconstruction using Generative Adversarial Networks and Hardware Performance Testing

This section presents a study into the viability of using Generative Adversarial Networks (GANs) as part of a particle physics hit reconstruction stack, culminating in the evaluation of the final algorithm on a range of hardware. The processor hardware tested included a new chip architecture, Graphcore's Intelligence Processing Unit (IPU). Firstly the motivation behind the holistic study is presented; a summary of track reconstruction techniques; an introduction to GANs; the simulation and datasets used in the study; the training process and results of the study; an overview of the hardware testing, the Graphcore IPU and results from the study.

6.1 Motivation

Data rates at large particle physics experiments like the LHC are in excess of 1 TB/s. To process the volume of data current experiments deploy large scale CPU farms, where the data rate is limited by the time taken to decide whether to store the event. Additionally an often underestimated cost is the computation of simulated events, the generating

and processing of which consumes the majority of the CERN compute budget. The current methods are insufficient for the projected computational cost of running future experiments, and as such increased interest in looking at alternative computing paradigms. These alternatives would replace CPU only compute farms with a mixture of CPUs and GPUs, and potentially Field Programmable Gate Arrays (FPGAs). The first level of the software trigger in the LHCb upgrade will run on GPUs [52] when it begins operation in 2021.

Historically particle physics has used tools like TMVA[43] for the types of multivariate analysis that was a precursor to the current wave of machine learning algorithms, however industry standard tools like CUDA [53] TensorFlow [32] and PyTorch [54] are being more widely adopted as the barrier to entry becomes lower and the potential performance gains become larger. Most particle physics analyses (including that in Chapter. 4) use ML methods as part of the final selection process, however neural networks are also found in particle identification [55], flavour tagging [56] and triggering [57, 58]. Motivated by the potential high performance of hardware accelerators like GPUs and FPGAs, novel methods for track reconstruction using machine learning have started to be developed. Alternative computing architectures like the IPU (Sec. 6.8.1) could potentially be of use in particle physics both in terms of performance gain and the flexibility offered by the architecture for alternatives to traditional HPC approaches.

Hardware improvements alone are not sufficient to handle the volume of data expected in future experiments, and attention is being paid to alternative algorithms to process the data. Of particular interest to the work presented in this chapter are Generative Adversarial Networks (GANs); these are pairs of neural networks that are trained simultaneously with competing objective functions. A discriminator network classifies the output of a generator network as either from a training set or fabricated by the generator, this information is used by the generator to inform its training [59]. Further details on GANs is given in Sec. 6.4. Outside of HEP GANs are typically used in generating image data from noise vectors, with extensions incorporating conditional information to direct the generation.

This has extended to conditioning a GAN with an input image resulting in a variety of style transfer [60] and resolution up-scaling [61] applications. Such results are broadly at research rather than application standard, however analysis quality data is becoming increasing possible to generate. These networks are now beginning to be used in HEP applications, to produce simulated data where the model is conditioned with physical information to avoid a rate limiting bottle-neck in traditional Mote Carlo approaches, and to model detectors. The proof-of-principle study presented in this chapter looks at using a GAN to correct detector level effects from particle hits and recover higher fidelity analysis quality data that can be fit with a track fitting algorithm. Current methods of track fitting such as the Kalman Filter in a simple form do not a priori model the physics of an interaction, and while extensions can be made to modify the Gaussian assumption of energy loss this can be computationally expensive. The simulation of particles through matter are well modelled in packages such as `GEANT4`[62, 63] and result in a specific energy loss function that includes all physics effects. If this information can be included during the reconstruction of hits in a detector it would be possible to more precisely account for detector level effects and improve reconstructed hit quality. Further details will be presented in Sec. 6.5.2. An evaluation of this algorithm across hardware choices will be presented in Sec. 6.8.

6.2 Track Reconstruction in HEP

Specific details of track reconstruction at the LHCb detector is given in Chapter. 3, however most particle physics detectors operate under similar general principles. Tracks are formed by a combination of hits on a sequence of detector planes that are connected with a fitting algorithm into a continuous track. Future liquid argon detectors based, such as the Deep Underground Neutrino Experiment (DUNE), observe a higher granularity of raw hits but are not fitted in the same way. Examples of both are shown in Fig. 6.1. In the results presented here the focus will be on tracks from detectors with discrete layers, however, in Sec. 6.5 it will become clear that the method is not limited to such tracks.

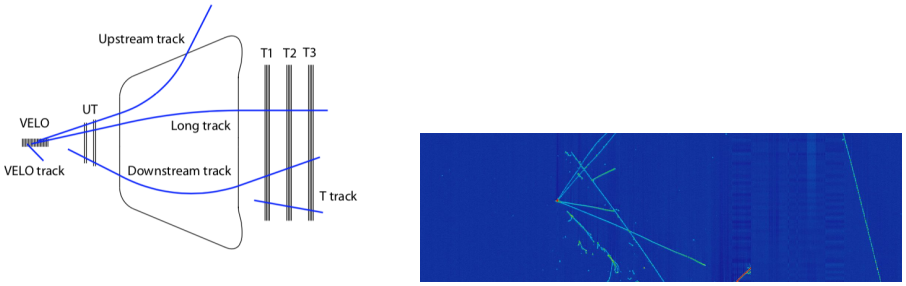

Figure 6.1: Picture of various track types observed in the LHCb detector (left) and example track from MicroBOONE neutrino experiment(right).

Linking observed hits in multiple layers of the detector to a continuous track requires a more sophisticated approach than brute force due to very large number of potential hit combinations that could form a track. The goal of fitting tracks is to obtain the most accurate estimate of parameters associated with a track. These parameters are then combined with PID and vertex information to determine properties of the particle such as invariant mass. The track is built from measurements, or track states, at each detector layer which combine a hit location with measurement noise, and a propagation direction. This direction is usually a tangent to the track segment in low B-field regions of the detector, but accounts for curvature from the B-field where appropriate. The track state is propagated between each hit using a transfer function combined with process noise that accounts for effects such as multiple scattering. The final track is therefore a continuous series of measurements with two sources of noise, both are assumed to have zero expectation value. These hits associated with a track are then processed using a Kalman filter to obtain a best fit to the measurements. A Kalman filter is a method of incorporating a series measurements and their uncertainties to best estimate the underlying real event, it is used across most of engineering and sensor applications, and in the LHCb experiment is used to obtain a smooth best fit to the particle hit locations. Key advantages of the Kalman Filter are that the measurements are incorporated one at a time, updating the track state at each measurement. The objective of the Kalman filter is to minimise the χ^2 of the fitted track, which is identical to minimising the least squares. However, the algorithm has the advantage that by proceeding iterativly it does not require large

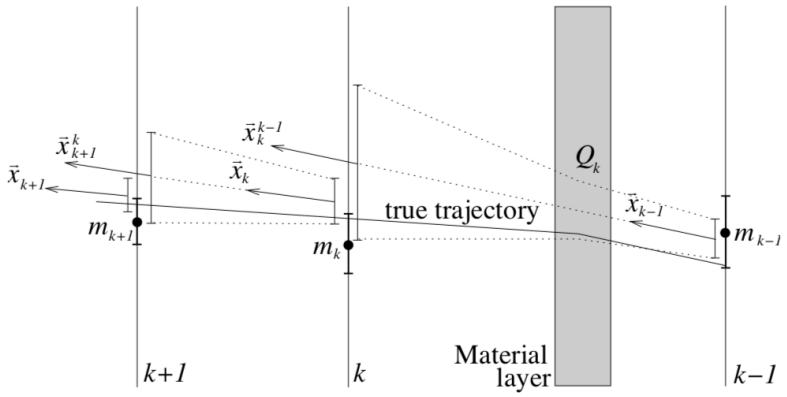

Figure 6.2: Diagram showing the iterative steps used in a Kalman fitter process propagating the measurement and uncertainty from layer $k-1$ to k and $k+1$. At each step the projected region is taken from the uncertainty of the individual hit and projected to the previous layer, as shown by the dashed region. Once all hits have been incorporated the smoothing functions result in a continuous true trajectory as shown by the solid line.

matrix inversions associated with a global least square approach that are computationally expensive and slow. This makes it an ideal solution for applications where low latency is important, either for real time responses of in order to manage a high data rate.

The Kalman filter first passes through the hits in reverse, then once all the data are incorporated, a forwards pass using a smoothing function results in the best fit result. The first stage of the Kalman filter implementation at LHCb is to take a position measurement with an uncertainty and predict back a region where the previous hit is expected to be. In Fig. 6.2 this is shown on the data point k-1 projecting the dashed region back to data point k. The filter function incorporates the next data point and updates the projection back towards the k-1 data-point. The projection is shown by the dashed lines, and contrasting the k^{th} and $k-1^{th}$ projections the impact of a material layer Q can be seen to introduce additional uncertainty. This iterative procedure continues until all the data have been included, however the resulting prediction may not be smooth, so the forwards pass applies the smoothing function on each section of the fit. The mathematical formalism can be found in Refs. [64, 65]. To pass the first stage of criteria as a successful track the Kalman filter must associate a set of hits as a continuous track, and the fit must pass a track quality criteria. This fairly stringent cut removes a very large number of

potential track candidates that do not meet the quality criteria.

This algorithm clearly relies on the quality of the data input to filter, if the uncertainties in reconstructed hits are reduced the quality of the track fit would increase. In turn if more tracks were of higher fit quality then a greater percentage of tracks would pass the initial selection and could potentially increase the track recovery efficiency.

In reality many particles undergo complex energy losses in their interaction with the detector. To model such losses like Bremstrahlung, the Kalman Filter needs to be modified to account for a non-Gaussian energy losses which can be computationally expensive. This motivates the study to look at alternative reconstruction processes that can incorporate a complex model of energy losses. The preliminary study presented in Sec. 6.5.2 looks at an alternative method to account for multiple scattering effects.

6.3 Basic Concepts for Neural Networks

Neural networks are a type of machine learning algorithm that is used to model a representation of data. They are most often used in HEP for classification problems in data, as implemented in Chapter 4. Using linear algebra neural networks transform an input to a desired output, where the exact function of the network is determined by the weights filling the matrices. For the weights in the matrices to be found, a "loss function" that quantifies the network performance is minimised, where the loss function is minimised the weights in the network will correspond to the desired output. In a classification task this might be the highest accuracy on the training set, as described by the loss function. Loss functions are typically more complex that simply stating the desired output, as by modifying the loss function more robust training can be achieved. An example of this can be seen in a categorical classification task, comparing the Mean Square Error (MSE)

$$\mathrm{MSE}(y, y') = \frac{1}{2}\sum_{i=0}^{n}(y_i - y'_i)^2 \tag{6.1}$$

to the Cross Entropy (CE)

$$\mathrm{CE}(y, y') = -\frac{1}{n}\sum_{i=0}^{n}(y_i \cdot \log(y'_i)), \tag{6.2}$$

where y is the target and y' is the predicted output. Minimising the MSE and the CE result in the same number of corrected classified samples, however the cross entropy has a more strongly diverging loss function which promotes training. This principle can be applied to any loss function for a neural network, the desired application is encoded in the loss function.

To update the weights in the network a partial derivative is defined linking the network weights with the loss function. This means that the weights can be updated to minimise the loss, which can be calculated using the chain rule. This process is done using gradient descent, and there are many optimisers that facilitate this.

The dataset is batched in smaller clumps of data for a mixture of efficiency, and promoting generalisation during training, as it is easy for a network to over specialise on training data and have degraded performance in application to real data. An epoch is defined by lopping through all the batches that make up the data, i.e. over one epoch in training the network sees all the data. These principles are extended beyond classification to generating data using a neural network in the next section.

6.4 GAN Theory

Generative Adversarial Networks are pairs of networks that, when trained in parallel, can be used to produce arbitrary data from an initial random noise vector. These are typically trained on image data using convolution neural networks, however the network structure is arbitrary and can be trained on any kind of data. One network, the Generator, takes an input and outputs some fabricated data, while the second network, the Discriminator, classifies the produced data as either from a real training set or fake. A picture of the pair of networks generating images from the MNIST images dataset is shown in Fig. 6.3. The discriminator, however, is not pre-trained, so it must reach a decision based on an un-supervised learning approach. In the initial phase of training the Generator will output data that very closely resemble random noise, these fake data are easily classified as such by the discriminator, reducing the loss function. For the generator component of the loss

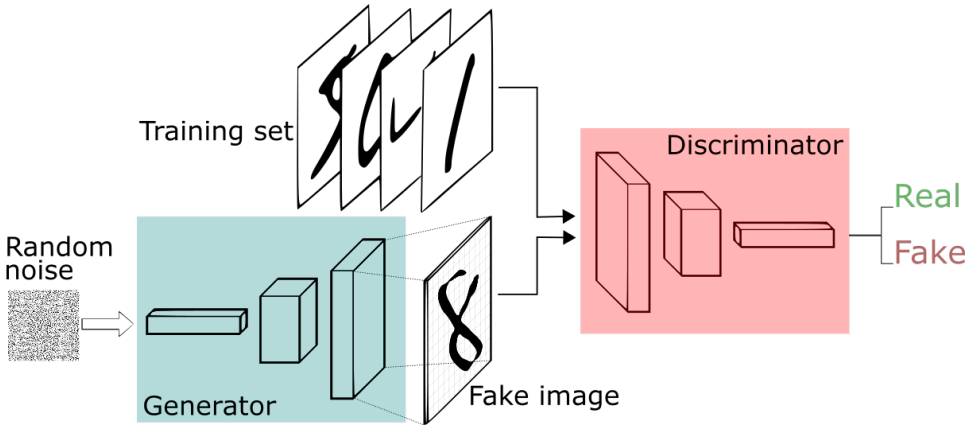

Figure 6.3: Simplified structure of a GAN where the Generator produces images (MNIST) from random noise in blue. These fake images and real training set images are classified with a Discriminator block in red which outputs a single decision as either real or fake.

function to be minimised the output data must start to resemble the true data to be incorrectly classified by the discriminator. In theory this process continues until the fake data are inseparable from the real data. The two networks are therefore a generator that outputs an image, and a discriminator that takes an image as input and that outputs a probability between 0 and 1 that the image is real or fake.

The competing contributions of the two networks can be formally defined in a loss function as follows:

$$L_{GAN} = \mathcal{E}_y \left[\log D(y)\right] + \mathcal{E}_z \left[\log(1 - D(G(z))\right], \tag{6.3}$$

which is the expectation value of the log of the discriminator classifying data y, and the expectation value of the log of 1 minus the discriminator output of the generator output based on random noise vector z. This exploits the range of outputs between 0 and 1 of the discriminator to be well defined and to encourage the desired performance. This loss function is turned into an objective function, $\mathrm{Q_{GAN}}$, by stipulating the competing goals for each network as minimising the value of L_{GAN} for the Generator and maximising the value of L_{GAN} for the Discriminator as:

$$\mathrm{Q_{GAN}} = \min_G \max_D L_{GAN}. \tag{6.4}$$

The competing objectives described in Eq. 6.4 are best dissected in two sections. Firstly, in order to maximise the objective function the discriminator must output probabilities

close to 1 for true data and close to zero for fake data as shown by the in the first term of Eq.6.3, and should output a probability close to 0 for the second term where the inputs come from the Generator. Secondly, as the Generator only appears in the second term of Eq. 6.3, to minimise the objective function the Generator must produce an input to the Discriminator that will yield a probability close to 1. Naively by minimising this loss function both networks will train together and the generated data will become inseparable from the true data.

In reality however, things are more complex. Firstly, a Generator often achieves images that are not visually satisfying, yet they are still sufficient to pass the discriminator. This represents the fundamental challenge. If the discriminator does not place weight on the same properties of the data that are desired, the Generator may align on a different component of the data. This is not surprising as the initialised networks know nothing of the images and are trained together. Secondly, the competing objectives in the loss function leads to a *Nash Equilibrium* [66]; where when there are two competing objective functions the optimal solution lies at a local saddle point, which is intrinsically unstable. Deviation from the saddle point during training often results in catastrophic failure that cannot be recovered from. Following on from [59], much of the research into advancing generative methods focuses on how to circumvent this Nash equilibrium, and avoid a phenomena know as mode collapse. Mode collapse is when a network ceases to train effectively after finding some extreme local minima in the loss surface, and outputs an essentially static image for all inputs.

The form of GANs discussed thus far are referred to as *vanilla*, which means they take random noise as an input and generate data based on that input. These vanilla GANs can by construction only generate data that is random, which has applications in HEP for faster Monte Carlo generation[67], however, for application to data manipulation is insufficient. An extension to the vanilla GAN model is that of a *conditional* GAN, where the input state is *conditioned* taking as input some nominal data that undergoes manipulation.

The simplest form of conditioning appends a meaningful label onto the random noise vector z used as input to the Generator. This when paired with the Discriminator allows for the generation of finely labelled data. A further degree of conditioning replaces the random noise vector z with an input dataset, typically an image, that forms the base of some kind of transformation. This is often a kind of style transfer where the desired content of an image is preserved and style is changed. A flexible example of style transfers across multiple domains can be seen in [68], and an example of up-scaling image resolution can be seen in [69].

6.5 eSRGAN Architecture

The architecture of the reconstruction algorithm eSRGAN is closely based on the Pix2Pix algorithm[68]. Pix2Pix is a GAN that performs image-to-image style transfers across a wide variety of domains, without need for extensive fine-tuning of the model or alteration to architecture. This means that the Pix2Pix algorithm can be used with virtually no modification in style transfer applications ranging from transforming satellite images to maps, segmented buildings into photographs, and digitally applying a DSLR focus on an image obtained from a smartphone camera. The core of this algorithm is detailed before the extensions applied for the reconstruction task.

6.5.1 Pix2Pix

The Generator architecture is relatively simple: a standard U-Net encoder-decoder design, which is shown in Fig. 6.4[70]. Such a network uses convolutional layers and down-sampling (reducing the size of an image with weights to predict pixel values) to reduce the size of the input image to find a restricted bottle neck or latent space, which is then inflated to the original image dimensions by up-sampling (when an image is scaled up, and the additional pixel values predicted) and using de-convolutional layers. The vanilla encoder-decoder structure forces all information through a bottle neck, highly restricting the representation

space, however, in image translation tasks much of the low level structure is consistent between input and output images. To preserve this low level structure within each layer of representation each of the up-sampling layers is concatenated with the symmetric down-sampling layer. This is advantageous as each layer tends to learn different degrees of abstraction, which can be used to guide the generation.

The discriminator is less typical in design; where traditional networks make a single decision on the entire image, the authors of Pix2Pix implemented an algorithm called PatchGAN [1] that provides an ensemble of decisions across the output image based on inputs of sub-regions of the output image. The PatchGAN architecture itself is a series of convolutional layers that outputs a single decision of a region of the image, typically 70×70 pixels. This serves to promote fine grain detail by only looking at local areas of the output image that would otherwise be missed if only considering the entire image. The PatchGAN is a type of Markov-based discriminator, in that is assumes independence of pixels separated by more than the sliding window width, which is a common assumption in a number of models describing texture of images. Therefore the authors claim that evaluating this discriminator can be thought of as a form of texture/style loss.

The loss function of scanning discriminator is paired with an L1 loss on the output image, which is a sum of the pixel wise difference across the real and fake images. The L1 loss ensures that the large scale structure has been preserved, and the use of such a rigorous additional constraint is only possible because the algorithm relies on pair-wise image pairs in training, if the generation was from a random noise vector this approach would not be possible.

The total loss function can be written as:

$$Q^* = \min_D \max_G \left[\mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_1(G) \right], \tag{6.5}$$

where λ scales the L1 pixel-wise loss. This flexibility makes it an excellent candidate as a base for the reconstruction study presented in Sec.6.5.2.

[1] A discriminator not a generator, for clarity.

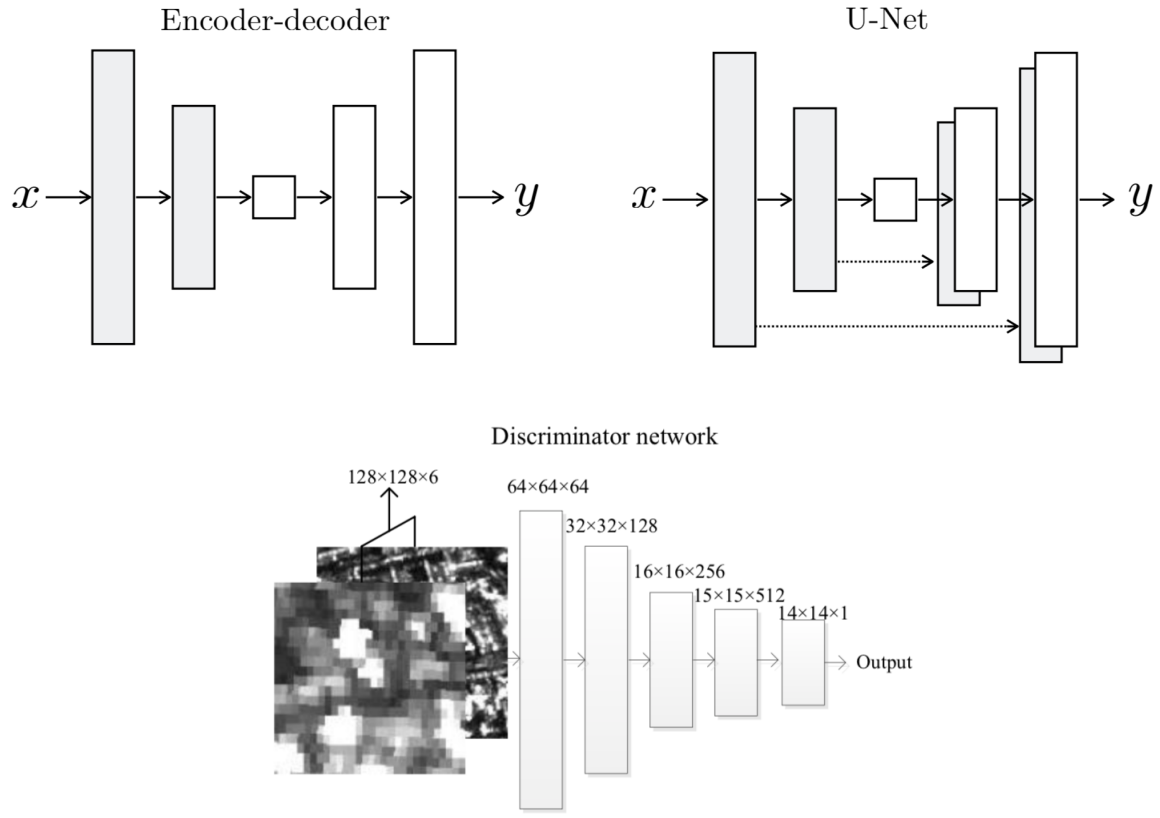

Figure 6.4: The U-Net structure used in the Pix2Pix algorithm is based of a standard symmetric encoder-decoder structure as shown on the top left, with the addition of the skip-layers where the feature map from each down-sampled layer is concatenated with the symmetric feature map from the up-sampled layers, as shown on the top right. The discriminator is a patch-GAN structure, which is scanned across the image to produce a spatial mapping of image quality. Each of the individual discriminators is a simple convolutions neural network with dimensions scaling as shown in the image. [68]

6.5.2 eSRGAN

The algorithm used here is referred to as eSRGAN, as reference to the original inspiration for the project. The algorithm is largely based on Pix2Pix, with a super resolution component added to increase the final image resolution by a factor of 2, and an additional L1 loss. The architecture of the eSRGAN algorithm is closely based on that employed to such success in Pix2Pix; where multiple applications and style transforms were shown across a wide range of domains. A U-net structure with concatenations between levels of feature maps is used as the generator and a patch-GAN is used as a discriminator, with

an additional CNN discriminator, and the L1 loss. When applied to a super resolution task there is an additional final de-convolution layer added to the generator. This has the additional challenge that no feature maps from the input can be concatenated onto this layer as the resolution is now greater than the input, but the simplicity of a single additional layer was favoured over a full residual block as found in [69].

6.6 Reconstruction

The aim of this investigation is to show that a GAN based reconstruction algorithm could be used in a future upgrade to the LHCb experiment (or future detectors in general) as a way to improve both reconstruction efficiency, and the quality, of reconstructed track hits. By improving the quality of reconstructed tracks hits there is a two fold improvement to data acquisition; firstly more tracks are recovered because their quality of fit is improved meaning that tracks that would previously be rejected could be recovered, but also that the improved quality of good tracks could improve the resolution of their physics parameters. The advantage of a GAN based method as an alternative to a traditional Kalman filter is that it would not rely on a simplified assumption of the processes involved in energy loss through particle interaction with matter in the detector. By using the experimental simulation, the full physics processes of energy loss due to particle propagation through matter is encoded in the Generator. The Generator then corrects the track hits prior to the track fitting algorithm finding the best estimate of the track path from the corrected hits. Extensions to the Kalman filter method can be employed to introduce similar non-Gaussian energy losses, however such methods tend to be computationally expensive. Therefore, a fast pre-processing stage prior to the traditional track fitting step that corrects for the effects of interactions with matter could be desirable.

This section will be split into two main sections: Firstly the simulation and construction of the toy data, then the training process of the algorithm.

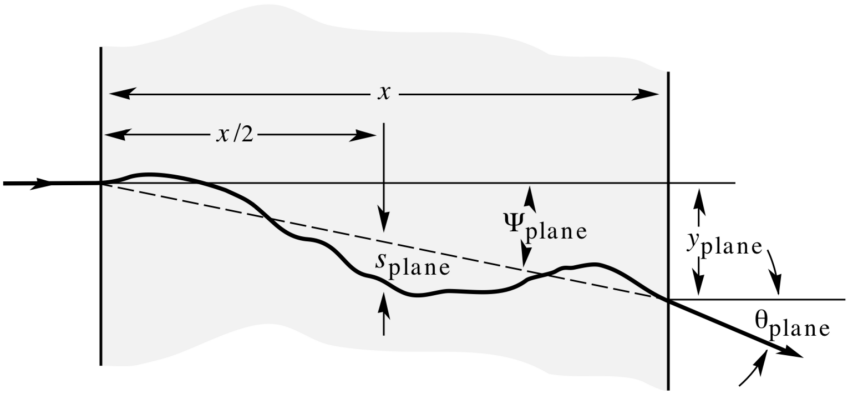

Figure 6.5: Schematic of Multiple Coulomb Scattering. The scattering of a charged particle through a layer of material can be described as a spatial shift and an angular shift, as such it is not necessary to model each scattering interaction in the material. The probability distribution describing these shifts is taken from Eq.6.6 and Eq.6.7. Figure from [37].

6.6.1 Simulation

In order to explore the potential of a GAN based reconstruction algorithm, a simple toy Monte Carlo simulation was constructed to generate a dataset of paired images to model physically meaningful transformations. Removing any explicit reliance on the LHCb software framework meant this study could be run in an efficient and lightweight manner to explore proof of principle applicability rather than specific implementation, and reduce the physics problem to the simplest feasible case. For simplicity, this simulation was limited to 2D to take advantage of faster computation and closer resemblance to the data used in existing algorithms shown in Sec. 6.5.1.

This preliminary study only focuses on the effects of Coulomb scattering. The Coulomb scattering process can be modelled using a probabilistic approach. The interaction of a charged particle through material causes the particle track to be continuously perturbed, a picture of this can be seen in Fig. 6.5 where a particle incident on a layer of material follows a complex path through the material. The scattering is considered purely geometrically, where any energy losses or decays within the material are ignored. The change to the path can therefore be described with two variables: θ_{plane}, the change in angle after traversing the material, and y_{plane}, the change in height after traversing the material. Following the

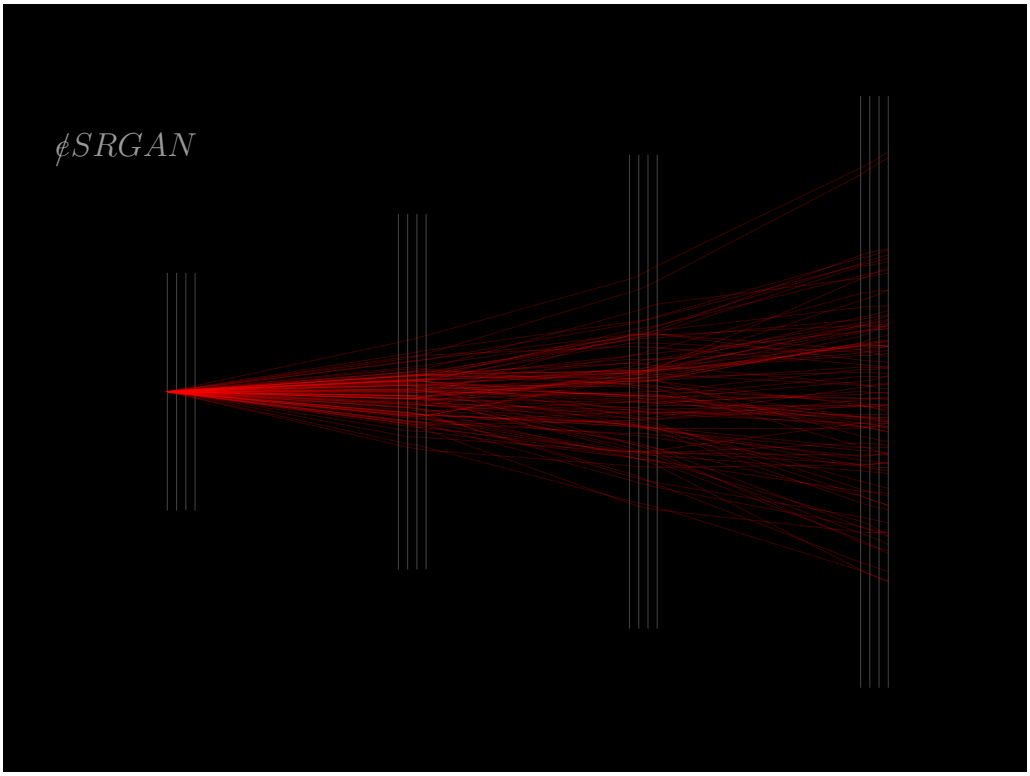

Figure 6.6: An example of several thousand electron tracks traversing through the toy detector, where all tracks originate from the same location with some initial variation in starting angle. Each track is propagated between each layer, where scattering effects are modelled, and the space between each hit in the detector layers linearly interpolated to connect the hits into a continuous track.

description in [37], these variables follow random distributions and can be described as:

$$\begin{aligned} y_{\text{plane}} &= (z_1 x\theta_0 1 - \rho_{y\theta}^2)^{1/2}/\sqrt{3} + z_2\rho_{y\theta}x\theta_0/\sqrt{3}; \\ &= z_1 x\theta_0/\sqrt{12} + z_2 x\theta_0/2; \end{aligned} \tag{6.6}$$

$$\theta_{\text{plane}} = z_2\theta_0. \tag{6.7}$$

Where in Eq. 6.6 and Eq. 6.7 z_1 and z_2 are independent random normal distributions, x is the thickness of the material and θ_0 is a small angle, where the scattering follows a Gaussian distribution, $\rho_{y\theta}$ is the correlation coefficient. With these probability distributions to describe the traversal of charged particles through some material budget, the simulation can be written such that the interaction of a single particle with a layer can be described, and the track properties updated. At each layer of detector a random displacement to the track direction and height is obtained.

The geometry of the simulation consists of four sets of four tracking modules, each consisting of four tracking layers. Each module loosely resembles a similar arrangement found in the LHCb TT layers, but is intended to be highly generic. Each of the sixteen detector layers was modelled as a pure silicon layer of width 600 μm. In each module the layers are separated by 2 cm, and each module is separated by 44 cm, these dimensions are arbitrary. No magnetic field is modelled in this simulation. Each particle is generated with a random initial angle, the particles are propagated though the detector and at each intersection with layers of the detector a Coulomb scattering angle and displacement are calculated. These shifts are then applied to the particle trajectory and propagation continued.

The truth-level track is generated at the same time, where the truth-level track is propagated as a straight line encountering no interaction with the detector planes. These pairs of images form the base of the training set. In both cases the only reconstructed hits are at the detector layers, and a linear interpolation is used to fill the path between each layer. This setup is a highly simplified model, the detector dimensions, resolution, and scattering angles are reasonable on an order of magnitude level. The track data is represented as an image with pixel intensity representing multiplicity; this means most image pixels are binary 1 or 0, however, when tracks overlap the count in the overlapping pixels increases. Alternative approaches could be considered starting from track parameters, however this is beyond the scope of the study presented here.

6.6.2 Datasets

Datasets generally consisted of 10,000 pair-wise image samples. The datasets used in the training and development processes consisted of:

- **Single tracks** - Single tracks undergoing random coulomb scattering in each layer, simplest possible example where the result is a single unambiguous straight line.
- **Multiple Tracks** - Image plates containing up to seven tracks, but typically two - five tracks. Here overlap is expected, and sometimes co-linear tracks are seen which

pose a particular challenge to reconstruction. The upper limit of seven tracks was motivated because both training time and dataset size needed to be increased as images became more complex, and was beyond the scope of this study.

- **Blurred tracks** - Some tracks were smeared with a Gaussian function to soften the binary description of the image and model some uncertainty in the real data.

- **Varied track origin** - In most cases the tracks all start from the same origin, but in some experiments datasets with tracks from various origins are used. This emphasises generalisability but also introduces significantly more variation in a relatively small sample data set.

- **Masked regions** - The final set of tracks had a blind mask randomly placed over the image that set all values to 0 within the mask. This mimics missing data though poor reconstruction or tracks that would typically not be recovered at the first stage, and could represent a large improvement in recovery rate.

- **Super Resolution** - Several datasets were generated with samples that had a true image of twice the resolution of the input image. These are to demonstrate that the algorithm can reconstruct basic tracks in a super resolution capacity, where a factor of two hits would dramatically improve the quality of the track fit.

Some example training pairs of images are shown in Fig.6.7 where a variety of experimental cases are shown, even with highly simplified construction notable experimental problems already become apparent such as track crossing and overlap such that individual tracks cannot be resolved from their neighbours.

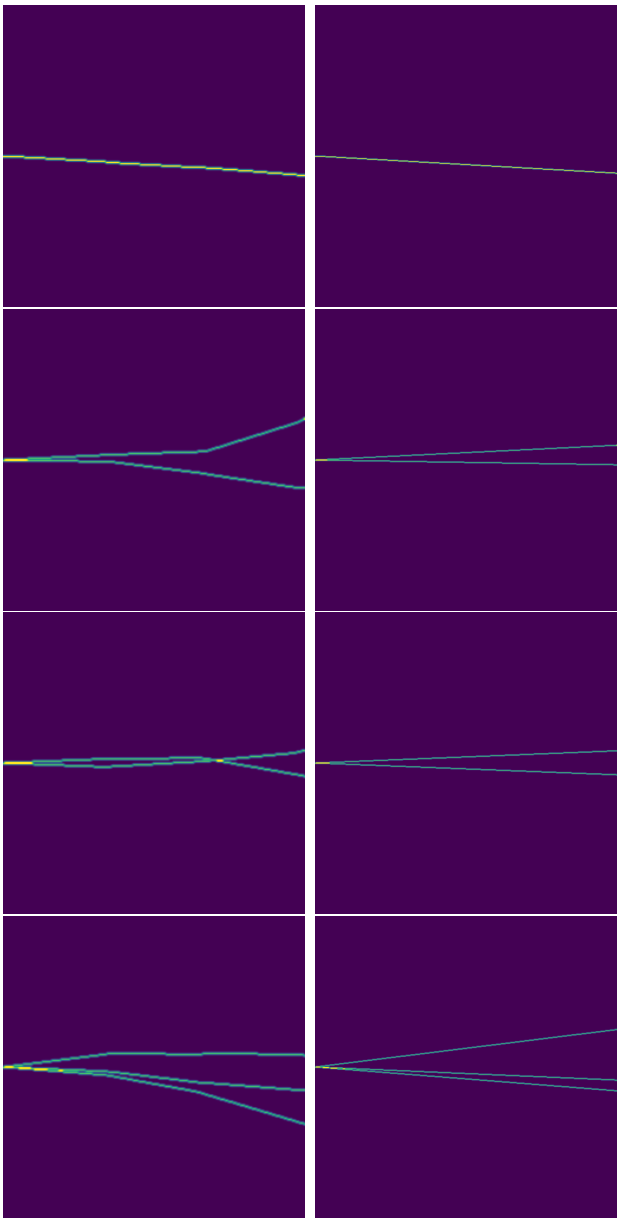

Figure 6.7: Training image pairs for the eSRGAN algorithm, demonstrating a variety of input conditions and the reconstructed target. The input images are shown on the left and the truth images shown on the right.

6.6.3 Training

Training of the neural network was done primarily on the `BlueCrystal` computing cluster to take advantage of Nvidia P100 GPUs[71], typically trained on a single instance rather than distributed across multiple cards due to the limited size of the model. Training on these GPUs compared to MacBook Pro CPU resulted in a factor of 10-20 times speed increase, significantly more flexibility with regards to training logistics, and allowed training to typically be run over a period of 12 - 24h hours.

Generally training datasets were approximately 10,000 image pairs, which was felt to give a good distribution of the core tracks, while not increasing the time taken to iterate through each epoch (a pass through the entire training dataset) of training exorbitantly. Obviously, as the dataset becomes more varied and more free parameters are introduced to the simulation the dataset size must be significantly increased to preserve good coverage over the core of the track distribution. For future work the exact composition of the dataset will become an important consideration, however given the scope of this early stage of investigation the size was deemed sufficient.

Training was generally run for 200-500 epochs depending on the task, and the loss monitored. It became easy to see when mode collapse had occurred as the loss would jump sharply and remain constant thereafter, and the output images remained unchanged. Mode collapse was more of a problem for this analysis than expected based on previous testing of the pix2pix algorithm, and ultimately was determined to be linked to the set up of the problem. Unlike typical style transfers using the `Pix2Pix` algorithm where the base structure remains constant; in the case presented here regularly the vast majority of the "on" pixels are going to change position, rendering direct translation of input structure less useful. Future work could look at alternative Generator architectures that do not expect as much of the structure of the image to be unchanged, or simply remove the concatenation layers in the U-Net structure. However it is possible that this structure encourages the correct number of tracks.

Evaluation of the performance of a generative network is complex. Various metrics can

be employed that quantify the noise level in a generated image, or how well a generated image performs against a pre-trained classifier, resulting in a classification score that quantifies both quality and diversity of results[72]. However, by far the simplest approach is to observe the output images for a visually satisfying result. While this is purely qualitative, for the purposes of this study visually satisfying images corresponded to a well behaved training process. Future work would benefit from quantifying the quality of the resultant track fit estimate from corrected hits as an unbiased way to measure the improvement.

6.7 Results

The simplest case shows a single track undergoing Coulomb scattering in each layer of the detector. This was performed with and without Gaussian smearing, and achieved similar quality results in each case. These tracks all originated from the same location to reduce complexity in the distribution to be modelled. Shown in Fig.6.8 is an example result when trained to recover a true image where the input track has resolution effects modelled using a Gaussian convolution. The GAN suitably recovers the straight true physics track.

The next set of results look at a detector plate with multiple tracks. In Fig. 6.9 and Fig. 6.10 exmaples with two tracks are shown. Given both diverging or crossing tracks in the input images the recovered set of tracks are very close visual matches to the truth. These results show non-trivial complexities that are vital to real applications. The multiplicity in detector pixels is a crucial component of resolving individual tracks.

The next set of results show a more encompassing dataset, by simulating data plates of up to seven tracks, and incorporating a random block masking to represent missing data that would require in-painting, this was designed to stress the simple model. However, as the results in Fig. 6.11, Fig. 6.12, Fig. 6.13 show, the GAN maintains strong performance. Slight degradation starts to become apparent in some cases. However this is not surprising given a training time of approximately 24 hours and a sample size of only 10,000 images. With more tracks and more variables the space of track corrections to be modelled becomes

significantly larger than in the two or three track case. Specific features of the core results to be highlighted are:

- In general the correct number of tracks are recovered, with only some difficult encountered with colinear tracks to model the difference between one or two overlapping instances.

- The reconstructed hits form a straight line for all tracks. This gives confidence that the Generator has encoded the true physics information from the unperturbed dataset.

- Large masked regions present minimal difficulty, with no alterations to the training procedure required and masks representing a moderate fraction of the dataset. These masks represent how an algorithm might fare against missing regions of data or short partially recovered tracks.

- Divergent tracks are usually well recovered even when there are multiple kinks in a track, suggesting that even in complex environments the GAN learns that kinks in a track do not represent the underlying physics case. This is seen strongly in Fig. 6.12.

- Finally, the robustness to a relatively co-linear column of tracks despite an initial high degree of degradation was found to be impressive, albeit pushing the limits of the resolution. However, the magnitude of the co-linear structure was broadly recovered effectively. This is seen to most clearly in Fig. 6.13.

The super resolution dataset takes an alternative approach, as in this case the architecture requires an additional de-convolution layer. This not only increased the number of parameters to learn but also does not benefit from the guidance of the concatenation layers in the U-Net structure. As such the dataset was simplified to input images containing three tracks or fewer, with no masks applied. As expected these models proved difficult to train, and suffered more extensively from mode collapse than in the previous cases. The

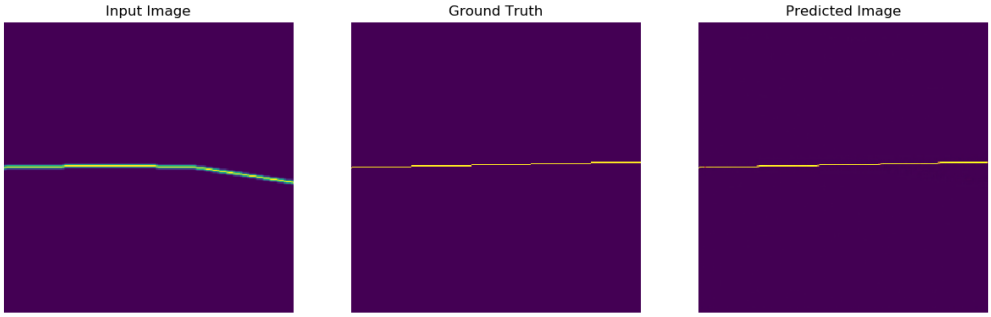

Figure 6.8: Single track recovery - Input track undergoing scattering (left), un-perturbed track experiencing no scattering (centre), generated true track produced by the GAN (right). In this case the single large kink is clearly removed and the true trajectory recovered, and the resolution effects are removed resulting in a clean binary image. This could have a substantial impact on the hit uncertainty in the detector through this recovery alone.

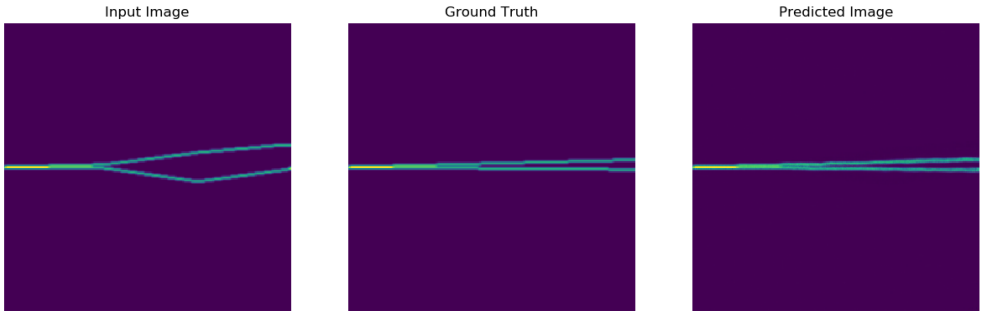

Figure 6.9: Clean Double lines - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). This simple case shows two tracks with a tight opening angle that separate and then undergo additional scattering effects. The propagation of the tight opening angle between the tracks is well recovered.

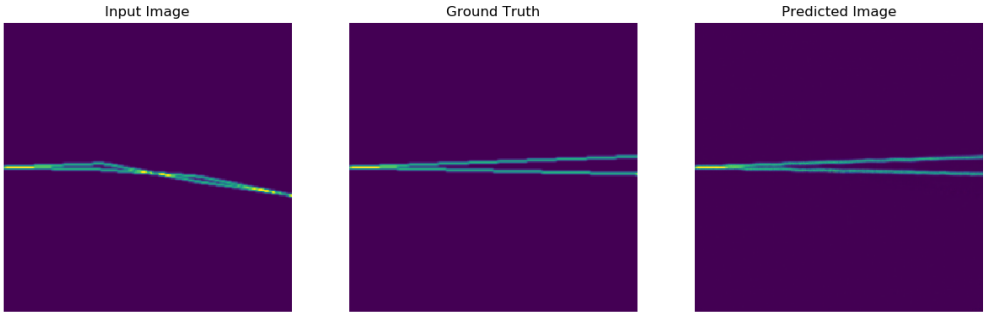

Figure 6.10: complex Double lines - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). The more complex input shows two tightly focused tracks that scatter such that they cross, and then cross again. This image looks much more like a single track in many ways than two distinct tracks, however despite this complexity the underlying tracks are well recovered.

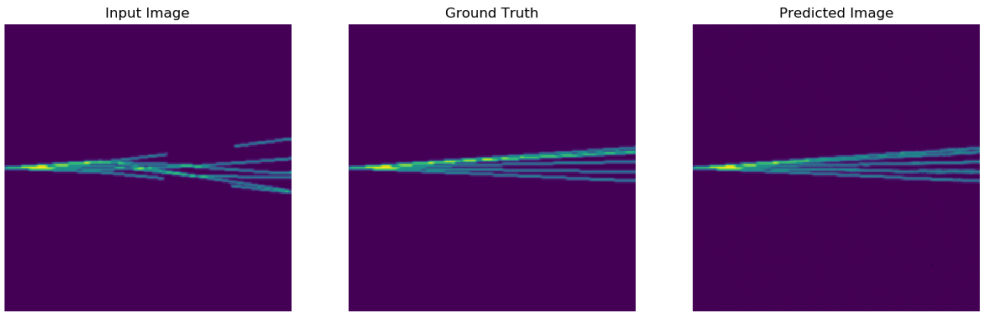

Figure 6.11: Six input tracks with masked regions on the input - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). The obvious complexity in this input image is the number of tracks, which both results in overlap near the origin such that the individual tracks cannot be resolved, and a complex multiple set of crossing occurrences. The masks mimic areas of the detector where data is missing. Despite the complexity, the correct orientation of tracks is recovered, with the exception of exact intensity on the upper doublet of tracks, suggesting one track has been missed from this input.

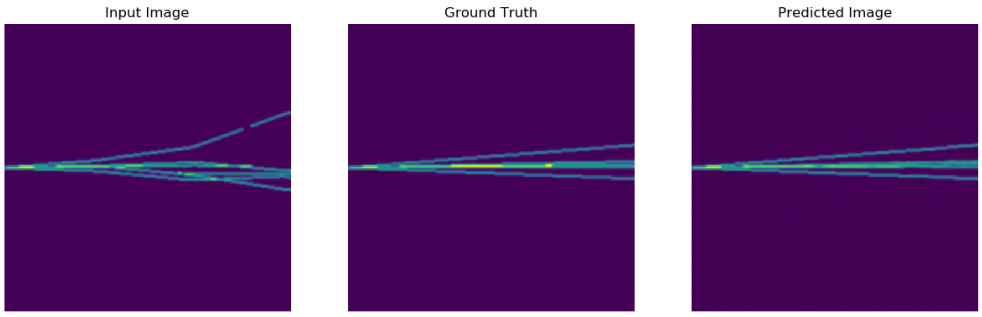

Figure 6.12: Five input tracks with masked regions on the input - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). This example is similar to that in Fig.6.11 but with a more extreme collimation of the tracks in the centre of the truth image. Here the spatial distribution is recovered and the intensity more closely matched.

results are shown in Fig.6.15 to Fig.6.16 where the visual similarity to the truth tracks is strong, and the scale on each image shown in pixels to make the up-sampling more readily apparent. These results are indicative of the power of the method explored here, information that is missing in the input image due to the size of image can be infilled correctly to recover a true track at 2x up-scaling. While the images are not substantially different than the results shown above, this indicates that the Generator has meaningfully encoded the physics information from the simulation and is able to apply this to achieve high fidelity results.

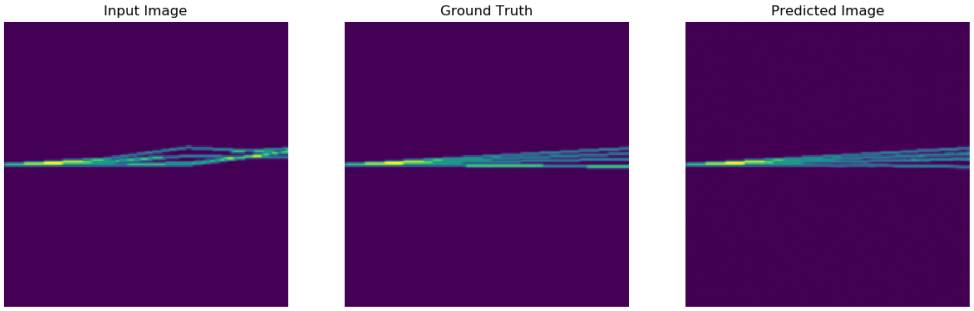

Figure 6.13: Four input tracks with re-convergence - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). These tracks are both tightly focused and undergo scattering that further reduce the separation, these tracks as an input are clearly difficult to separate, however, the GAN is able to recover the fine grain detail separation between each track and produce a tightly focused set of tracks.

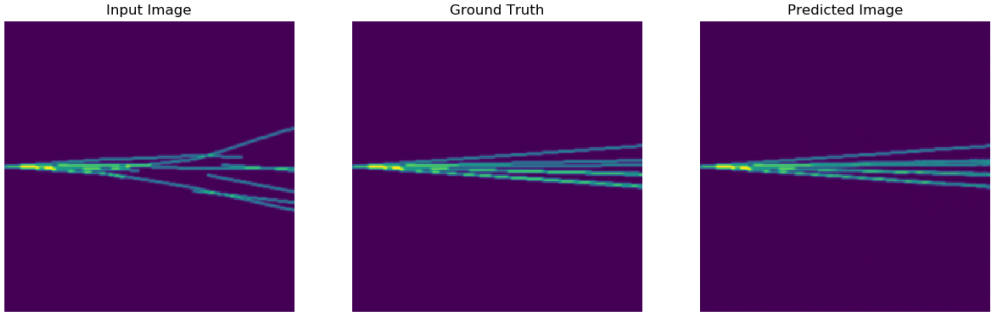

Figure 6.14: Seven input tracks with large scattering angles and significant masked regions - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). This represents one of the more complex environments that it was possible to recover, and represents a scenario that would be very poorly recovered using a standard fitting procedure. The truth image is complex given the large number of tracks, with two pairs of overlapping tracks, a pair of close tracks and one single track. The fidelity of the generated image is slightly less that in some of the simpler examples shown, but has each of the features of the truth track correctly reproduced.

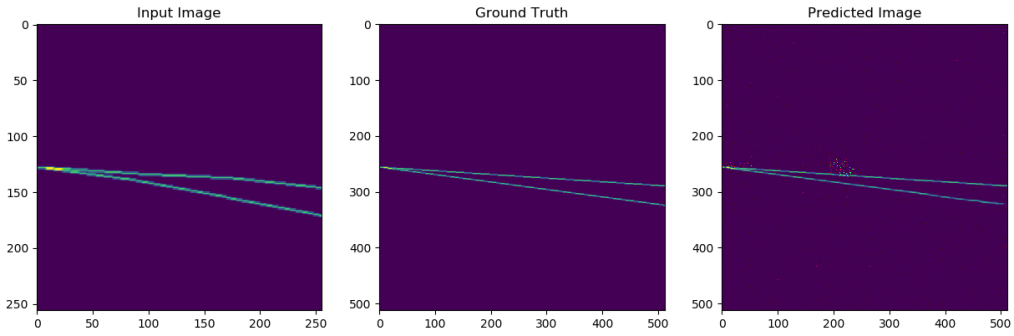

Figure 6.15: Super Resolution Recovery - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right).

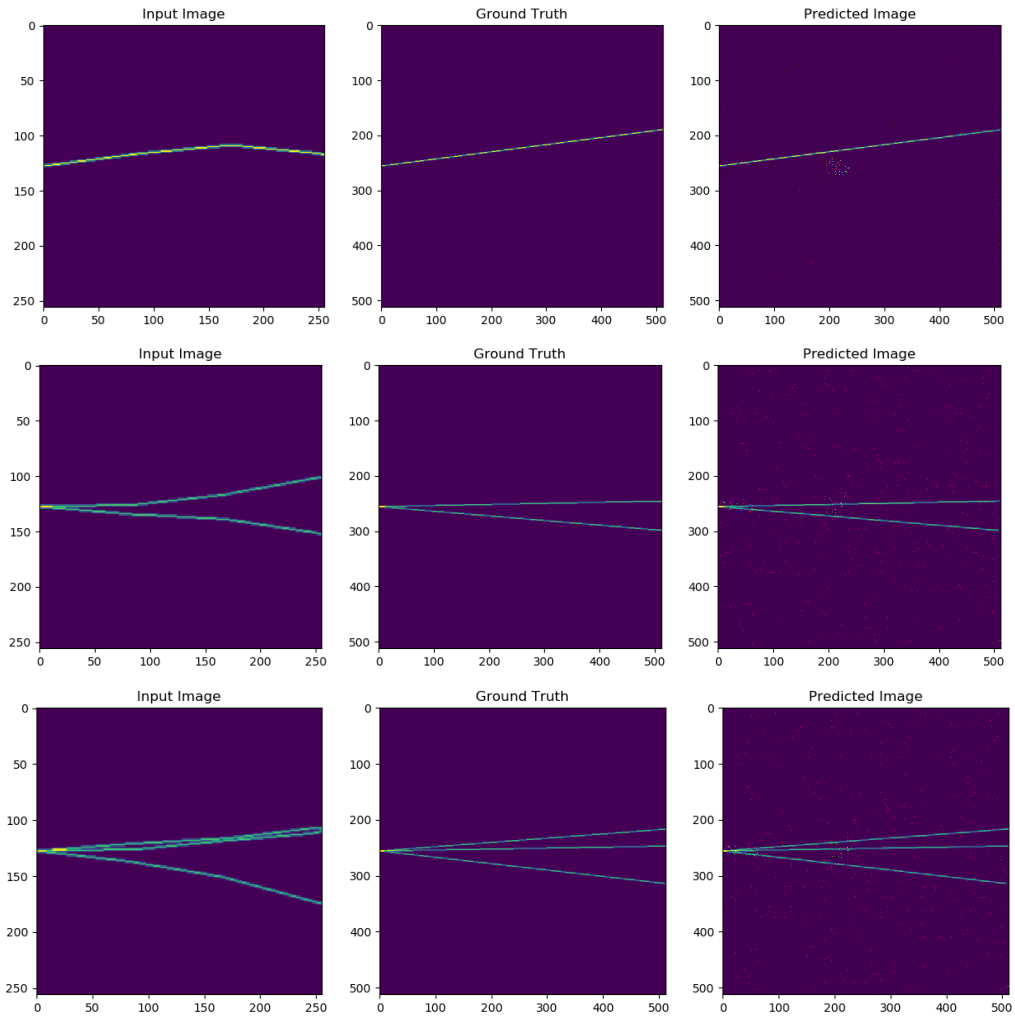

Figure 6.16: Super Resolution Recovery - Input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). In all instances shown the GAN recovers the high resolution detail of the hits, however, some noise artefacts are present in the generated output.

As a final set of results the training was extended to an environment that was expected to prove too complex for the relatively simple model and short training time. In addition to the multiple tracks of up to seven candidates with masks randomly applied, a variation of initial starting location was introduced as well. The same training conditions were applied as to the results without the shifted initial value, the results are shown in Fig. 6.17. Clearly these are of lower quality and recover the tracks with less reliability than any previous cases shown, but given the training conditions this is impressive, as straight line distinct tracks are generally understood to be the target, the precise mapping from input orientation to ground truth needs to be finessed. This was the only dataset where the

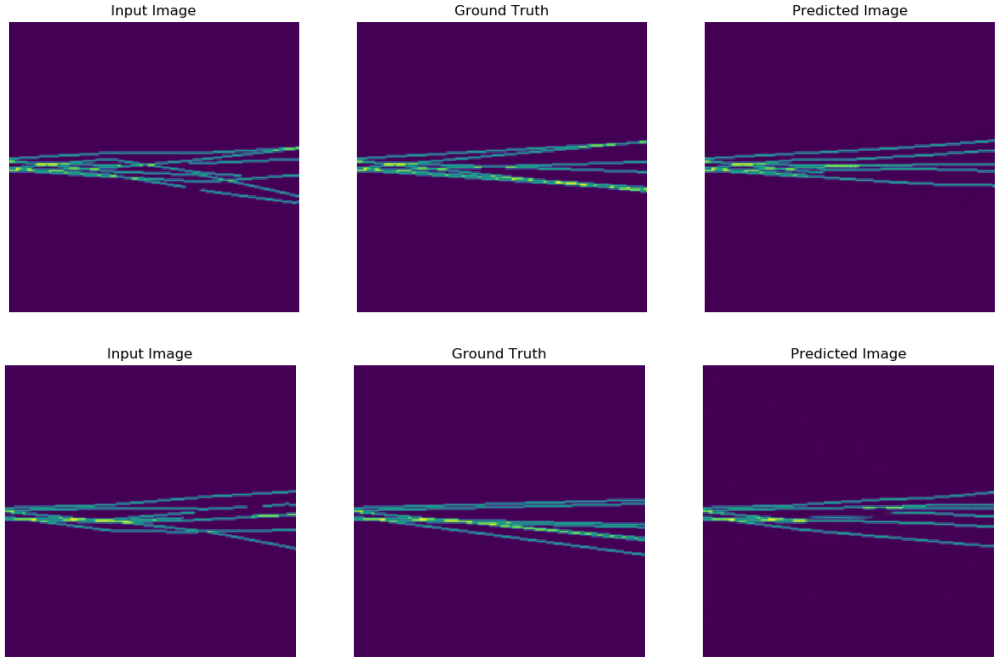

Figure 6.17: Multiple tracks with a spread of origin points - Two examples shown with input tracks undergoing scattering (left), un-perturbed tracks experiencing no scattering (centre), generated true tracks produced by the GAN (right). In these cases the GAN was not capable of recovering the all fine grain structure based on the variable input coordinates, however, the structure of the output images shows good initial encoding of track objects.

Generator did not learn that all tracks should be straight. Provided with a longer training run these results may improve. Even in this case where the results start to break down interesting features become apparent and insight can be gained into the training process that ought to aid development going forward.

- The learning propagates from left to right as the model is trained, this may be a feature of the general broadening and smaller difference between tracks on the left than the right, or may show that the model learns it squashed the region of uncertainty.

- With multiple starting locations the model begins to struggle to differentiate between tracks on an angle from a displaced PV and a scattered track at a later stage of the simulation.

- Including a variation in the track origin increases the potential space of example

tracks dramatically and with a dataset of 10,000 images is in-sufficient to provide good coverage of the full range of potential tracks.

These results show promise that detector level effects could be removed from reconstructed hits using a GAN in HEP experiments. Further work is needed to extend the modelling to an environment more representative of a real detector with specific physics requirements. This will result in a significant increase in training load and would expect to need to distribute the training over multiple devices to obtain a reasonable time to train.

6.8 Hardware Testing

As the algorithms used in HEP become more complex, both in terms of deploying deep learning models and more traditional HPC processes, the demand for increased compute power becomes paramount. This is a problem not unique to HEP and in industry there has been a strong development of dedicated hardware to aid high level computational problems. The processor types that could be used to can be categorised as:

- **CPU** - Traditional CPUs are highly versatile and have decades of dedicated high performance libraries that perform efficient computation. However, versatility generally limits performance.

- **GPU** - The rise of deep learning has dramatically pushed forwards the performance of GPUs over the last decade. They are specialised matrix multiplication machines, as these are the kinds of problems that require solutions in graphical rendering. Good GPUs generally offer an order of magnitude performance gain over standard CPUs for machine learning problems. Increasingly machine learning problems look sparse and the overhead to maintain performance balloons.

- **FPGA** - Allow for customisation of the chip for specific application needs. They can often provide extremely high performance but present a significant barrier to entry in terms of expertise, and while can be almost infinitely customised are not flexible for application switching easily. They may have more importance as part of the compute for a dedicated detector, but are not so general purpose.

- **Novel** - Several companies are developing novel hardware architecture to efficiently solve future ML problems. Notably Google have developed a Tensor Processing Unit (TPU)[73], Cerebras produced the worlds largest single silicon wafer dedicated to AI compute[74], and Graphcore have developed an Intelligence Processing Unit (IPU)[75].

As a local company Graphcore was approached by members of the Bristol LHCb group for a proof of principle study to evaluate the performance of IPUs in a HEP context. The purpose of the study was to explore the applicability of IPUs in a future upgrade to firstly the LHCb experiment, but also to evaluate the hardware for use in HEP more generally. The eSRGAN algorithm was deployed onto the IPU and a series of experiments conducted to evaluate the performance over CPU, GPU, and IPU hardware. This work formed part of a larger study evaluating whether other machine learning methods, as well as traditional HPC workloads, could be efficiently deployed onto the IPU for use in HEP. The findings of the study are published as a pre-print [76] and submitted to a journal. What follows is a summary of the relevant details pertinent to the authors contribution.

6.8.1 Graphcore's IPU

Graphcore's Intelligence Processing Unit (IPU) is a new design of processor, purpose built for modern machine learning algorithms and exploits a massively parallel design. A limitation of GPUs in ML applications is their Single Instruction Multiple Data (SIMD) architecture that efficiently applies the same computation on vectorised data. While initially very efficient for neural network computation, modern algorithms are increasingly utilising more complex interconnected and sparse structures. The Multiple Instructions Multiple Data (MIMD) architecture of the IPU allows a huge number of parallel threads to be run on sparsely accessed data distributed across the chip in the Static Random Access Memory (SRAM) built into each processor tile. This allows a more flexible, and potentially more efficient, utilisation of compute resources.

The hardware used in the studies presented here is Graphcore's first generation IPU, the Colossus™ MK1 GC2 IPU. The IPU consists of 1,216 individual processor cores each with 256 kB of local memory, with each IPU capable of executing a total of 7,296 threads in parallel. Each IPU has 300 Mb of memory in processor. A pair of IPUs is mounted on each card. A diagram of the IPU is shown in Fig. 6.18 which shows the two IPUs on each card, with the distribution of processor tiles and SRAM show.

Table 6.1: Key specifications for the hardware used in this study [71, 75, 77–79]. Performance in terms of floating point operations per second (FLOPS) is given for 32 bit single-precision. Thermal Design Power (TDP) is given for each processor, where for the IPU this is half of the total board TDP.

	Name	Cores	Memory	Clock Speed	TDP
CPU	Intel Xeon Platinum 8168	24	732 GiB	2.7 – 3.7 GHz	205 W
GPU	Nvidia Tesla P100	3584	16000 MiB	9.3 TFLOPS	250 W
IPU	Graphcore Colossus™ GC2	1216	286 MiB	31.1 TFLOPS	*120 W

IPU Exchanges™ link the tiles providing a high bandwidth and low latency communication channel of 7.7 Tb/s, and 80 IPU links connect the pair of IPUs with a total chip-to-chip bandwidth of 2.5 Tb/s. This link to external memory at a very high bandwidth means that the processor is not limted by the on-chip memory. To connect the IPU card to the host system 16 PCIe Gen4 links (8 per IPU) are used. The IPUs used in this study were mounted in a DELL DSS8440 IPU server with a total of 16 IPUs, and two Xeon Platinum 8168 CPUs with 24×32GB 2.4GHz DDR4 DIMM Modules.

The Graphcore software stack is provided in the Poplar Software Development Kit (SDK) that allows the IPUs to be accessed from popular ML frameworks such as TensorFlow, PyTorch, and Graphcore's own optimised framework PopART. Additionally the IPU can be accessed through POPLAR to directly program the IPU.

The drivers and SDK receive regular improvements through updates, however the results obtained in this study use SDK version v1.2.0.

The results presented in this study compare a single IPU against an Nvidia TESLA P100 GPU and a Intel Xeon Platinum 8168 CPU, where it should be noted that the IPU utilises approximately half the power of the GPU as only a single IPU was used at a time. Most commonly a full server of 16 IPUs are utilised, however, for the studies undertaken here a single IPU was used for simplicity.

Since conducting this study Graphcore released its second generation IPU, the Colossus MK2 C200 with 1,472 tiles and triple the local memory per tile.

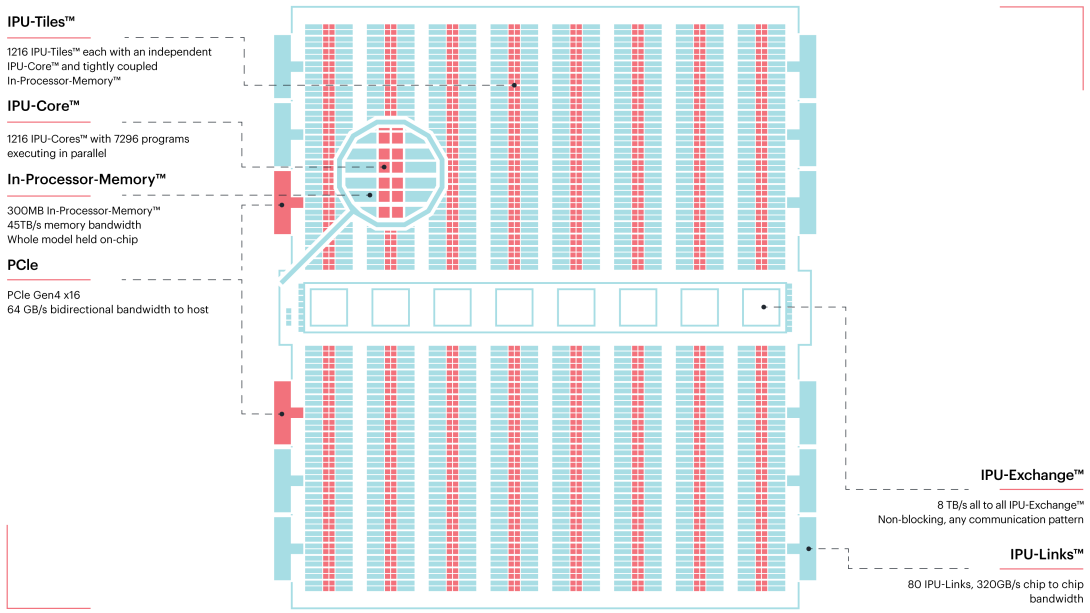

Figure 6.18: The Graphcore Colossus™ MK1 GC2 IPU[75]. The IPU-tiles™ are shown in red, and the In-Processor-Memory™ is shown in blue. The IPU-Exchange™ sits between the two IPU processors, the IPU-links™ allow the chip-to-chip communication, and the PCIe cards allow communication off chip.

6.8.2 Evaluating the eSRGAN Algorithm on the Graphocre IPU

A study was undertaken with Graphcore to evaluate potential use cases in HEP where the use of IPUs would be beneficial. This study included event generation using GANs, particle identification, and Kalman filters in addition to evaluating the eSRGAN algorithm. In this study the performance of a single first generation IPU is compared to an Nvidia TESLA P100 GPU and Intel Xeon Platinum 8168 CPU. The power consumption of the single IPU is approximately half that of the GPU. Key technical specifications of the IPUs, GPUs and CPUs used are given in Table. 6.1. The potential advantage of the IPU is firstly that as an architecture it has been developed to maximise performance for machine learning algorithms. However it would be reasonable to expect that applications taking advantage of the MIMD architecture that currently are not feasible with the SIMD architecture of GPUs.

The approach to training machine learning algorithms differs slightly when deploying algorithms onto the IPU compared to a GPU or CPU. This is largely due to the smaller memory on a single IPU of only 286 MiB compared to typically tens of GiB on a GPU. As a result a machine learning application must be written with much more concern about memory usage. For a small model this typically means a significantly lower batch size than would be expected on a GPU. To fit larger larger models into memory the use of multiple IPUs is required. Graphcore's framework provides the capacity to pipeline a larger model over many IPUs, however for simplicity of the study this was not investigated as it introduces an additional source of uncertainty.

The reconstruction algorithm presented in Sec. 6.5.2 has $\mathcal{O}(10^6)$ parameters, which is too large to fit onto a single Mk1 IPU. Therefore, two simplifications were made to reduce the size of the algorithm and ensure the comparison is one of hardware not insufficient optimisation. Firstly, the algorithm was run for inference (generating images, not training) only to measure the generation rate rather than training the model as this requires a second network and further restricts the memory. Inference speed is the metric of interest for application as part of a reconstruction chain. Secondly, the input image resolution was decreased from 256x256 pixels to 64x64 pixels. This does not fundamentally change the nature of the algorithm as it only requires removing two down-sampling layers and two up-sampling layers, but significantly reduces the memory overhead of the model.

The generation rate was recorded for the algorithm for the three types of processor for a range of batch sizes from 1 image to 1000, the results of which are shown in Fig. 6.19. The IPU outperformed the other hardware options for this application, significantly outperforming the CPU across the full range of batch sizes, but also out performing the GPU across the full range. The plateau for batch sizes of greater than 100 samples is interesting as it indicates that despite the greater memory to load data into, the GPU was not able to utilise this capacity to provide a notable performance benefit. Due to the algorithm architecture being primarily convolutional layers, which the IPU is know to handle efficiently[80], this improvement was approximately as expected.

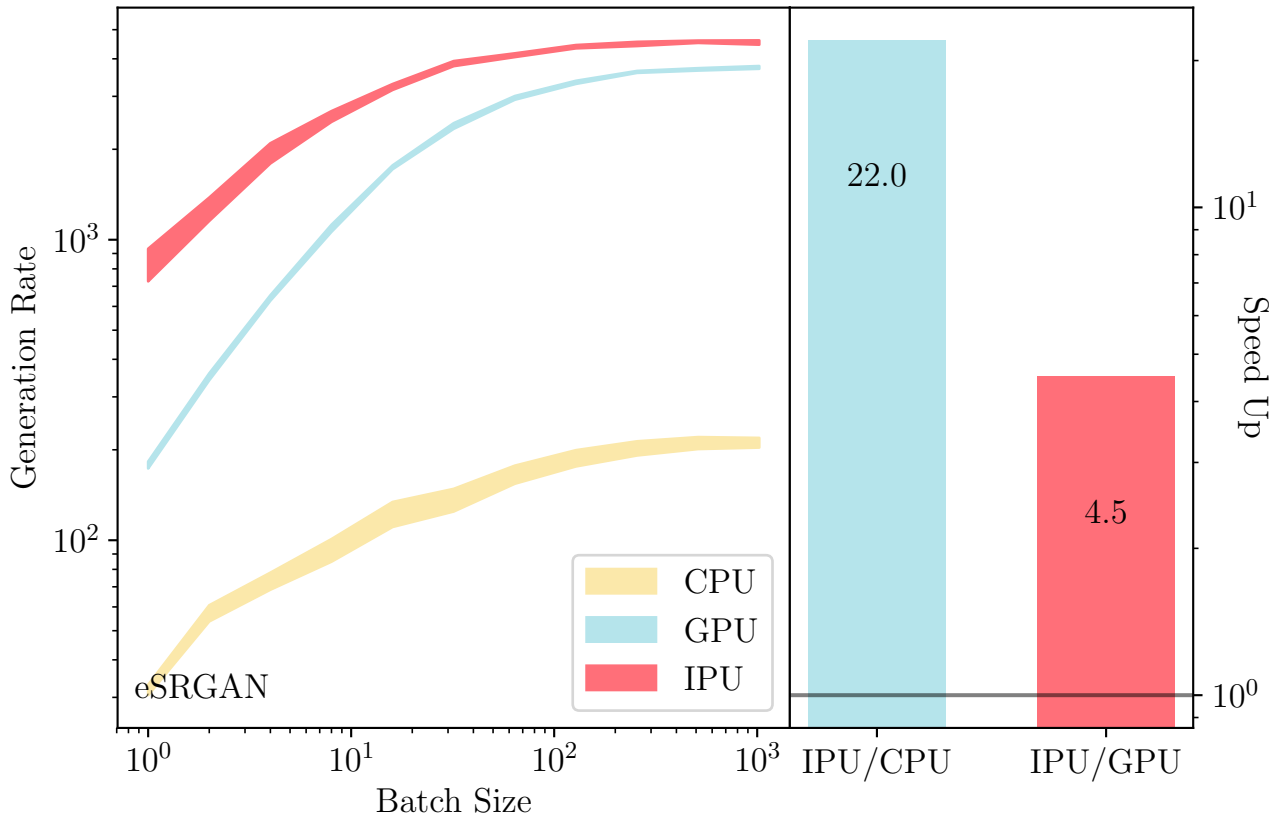

Figure 6.19: (Left) Generation rate of the eSRGAN algorithm for three types of processor hardware (detailed in Table. 6.1) shown as a function of batch size. The width of each line shows the uncertainty bands for each result. (Right) The ratio of speed up obtained with the IPU against the CPU and GPU respectively. The IPU outperformed the GPU at all batch sizes, achieving a peak improvement in generation rate of 4.5× at a batch size of 1.

The key region of performance benefit was for smaller batch sizes where the IPU most strongly outperformed the other hardware choices, reaching a maximum advantage at a batch size of 1 sample. The ratio comparisons shown in the right hand plot of Fig. 6.19 are taken for a batch size of 1, as this is the design batch size of the algorithm. At this batch size the generation rate of IPU is 22 times higher than the CPU, and 4.5 times higher than the GPU.

6.9 Future Work and Conclusions

The eSRGAN algorithm detailed here shows the development of a novel hit reconstruction method based on selected hits from a detector, that can recover an unperturbed track

with Coulomb scattering effects removed. The model encodes information about the translation from reconstructed hits undergoing scattering to a true unperturbed case. As a proof of principle concept, this provides a benchmark for algorithm development going forward. This study provides a qualitative demonstration that improved detector hits that can be recovered by encoding scattering information into a GAN, further qualitative work is needed to estimate the performance gain achieved with such a method. Specific developments for this algorithm need to consider more realistic applications to experiment, in terms of both geometry, and physical interaction. The addition of bremsstrahlung and picking a known detector silicon strip geometry would show exactly what defects could be recovered. This may be in the exact geometry of an experiment like LHCb, or looking at the wider field of particle physics experiments that will be online in the coming years as a potential method to deploy reconstruction. Additionally once the LHCb geometry has been incorporated into the simulation a direct comparison of physical parameters as a result of the Kalman fitter will give a benchmark on the improved resolution and recovery rates.

The hardware evaluated shows that there is a strong potential use case for IPUs in HEP applications, where the low latency and high throughput advantages are beneficial. From a hardware point of view it would be interesting to see how the performance scales over larger models utilising pipe-lining over multiple IPUs. Done efficiently, pipe-lining can increase throughput as by linking multiple IPUs to perform separate stages of the computation and passing the data through in sequence ensures that all processors are used to maximum efficiency, and reduces the I/O overheads.

Chapter 7

Conclusion

The work presented in this thesis covers three main areas: a first-of-its-kind angular analysis of $B^+ \to K^+e^+e^-$ decays, using data gathered at the LHCb experiment from 2011 to 2016; the upgrade of the RICH mirror alignment system to run online during data taking, and the testing of RICH mirror prototypes in the lab; and a proof-of-principle study into a novel GAN-based hit correction algorithm for future HEP experiments.

In the $b \to s\ell\ell$, sector there have been recent observations that point towards possible LFU violating effects in (axial-)vector couplings. To further probe the LFU assumptions in other couplings, precise measurements of the SM are vitally important. This angular analysis follows closely the work in the R_K analysis [4], sharing the samples and selection procedure. This provides a common ground for validation cross checks and important context for the result of the angular analysis. This angular analysis provides a stringent constraint on NP contributions through (pseudo-)scalar and (pseudo-)tensor couplings in the electron mode. There is only sensitivity to (axial-)vector couplings if large (pseudo-)scalar and (pseudo-)tensor couplings from NP are present for $q^2 \sim m_e^2$. As a result, this analysis can provide vital understanding of the electron reconstruction efficiency in $b \to sll$ transitions assuming no new physics in (pseudo-)scalar and (pseudo-)tensor contributions.

Events are selected, fine-tuned corrections to simulated samples are obtained, and the reconstructed mass distribution is fit for $B^+ \to K^+e^+e^-$ decays. Using the yields obtained from the mass fit, the backgrounds are subtracted from the angular distribution, and the resulting data unfolded into a $2-5\,\mathrm{GeV}^2$ bin of true q^2. The acceptance efficiency is modelled, and the angular observables A_{FB} and F_H are determined in a single binned likelihood fit to the angular distribution. The statistical uncertainty is evaluated using a two-dimensional Feldman Cousins construct, and the angular observables are found to be consistent with the SM predictions within the 1σ confidence interval. This work is presented in Chapter.4.

Particle identification is of vital importance at the LHCb experiment in all physics analyses, and is primarily done using the RICH detectors. Between data taking in 2016 and 2017, the alignment and calibration of the RICH detectors was moved online so that the alignment could be updated in real time during data taking. As part of this thesis, thresholds for updating the alignment constants were determined and the system for providing the alignment constants incorporated into the real-time LHCb framework. Additionally, as part of this thesis, mirror prototypes for the upgrade of the RICH1 detector were tested, and the framework for analysing the mirrors by evaluating the $D0$ spot size was developed. The author's contribution to both areas of work are given in Chapter. 5.

A proof-of-principle study that explores the potential use case of GANs in particle hit reconstruction at HEP experiments is presented in Chapter 6. Particle interaction with matter is well modelled in simulation, and if this full physics information could be exploited during the reconstruction phase there could be benefits to the resulting track quality. By using a GAN, it is possible to encode nuanced information about how particles lose energy as they interact with material in the detector, and this could offer an avenue towards an alternative reconstruction approach. Using a GAN based on the Pix2Pix algorithm, Coulomb scattering effects were removed from simulated tracks caused by their interaction with a lightweight generic detector. This method showed efficacy

over a range of conditions, with up to seven concurrent tracks and a large degree of data degradation. Additionally, it was shown that this method can be extended to increase the resolution of the generated track by a factor of two. This algorithm was deployed onto the Graphcore IPU™ as part of a study into the viability of alternative hardware choices for the next generation of HEP experiments. The IPU outperformed benchmark CPU and GPU throughput when generating correct hits with the GAN, and obtained a peak performance advantage at low batch sizes. This indicates that synergistic algorithm and hardware choices could be a viable avenue to explore in improving future reconstruction approaches.

There is a consistent picture emerging to describe the anomalies in the $b \to s\ell^+\ell^-$sector, with strong hints of potential NP. The LHCb collaboration will continue to probe NP in this sector over the coming years, and further updates to LFU tests, and angular analyses are expected. This angular analysis of $B^+ \to K^+e^+e^-$ decays contributes to these efforts.

Appendix A

Appendix

A.1 Corrections to simulation

The input of the residual corrections applied to the simulation samples as detailed in Sec. 4.5 can be found in the following figures. They show firstly the residual disagreement between fully selected $B^+ \to J/\Psi(\to e^+e^-)K^+$ data and simulation samples with a selection on the constrained mass and the original weights applied, then show the improved agreement once the original kinematic weights are combined with the additional multivariate weights. These final samples using the combination of weights are representative of the simulation samples used in all further stages of this analysis. Plots are shown for each trigger category for both runs such that the full effect of weighting each distribution can be seen. The magnitude of correction weights is small. In Run1 0.026% of the events have a weight larger than 4, and in Run2 0.032% of the events have a weight larger than 4.

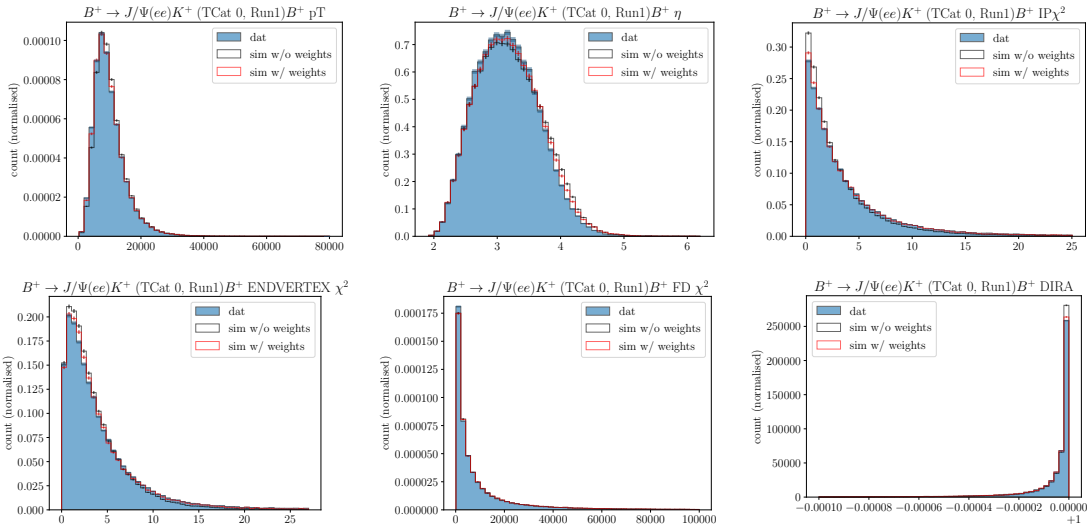

Figure A.1: The mis-modelling between the fully selected (including a cut on the constrained mass) simulated control mode samples and data is shown in the six input variables to `GBReweighter`. The data is shown in blue, while the simulated samples without weights shown in black, and the same simulated samples with the `FinalWeights` applied shown in red. The agreement is clearly strongly improved across all six variables. Run1 - eTOS.

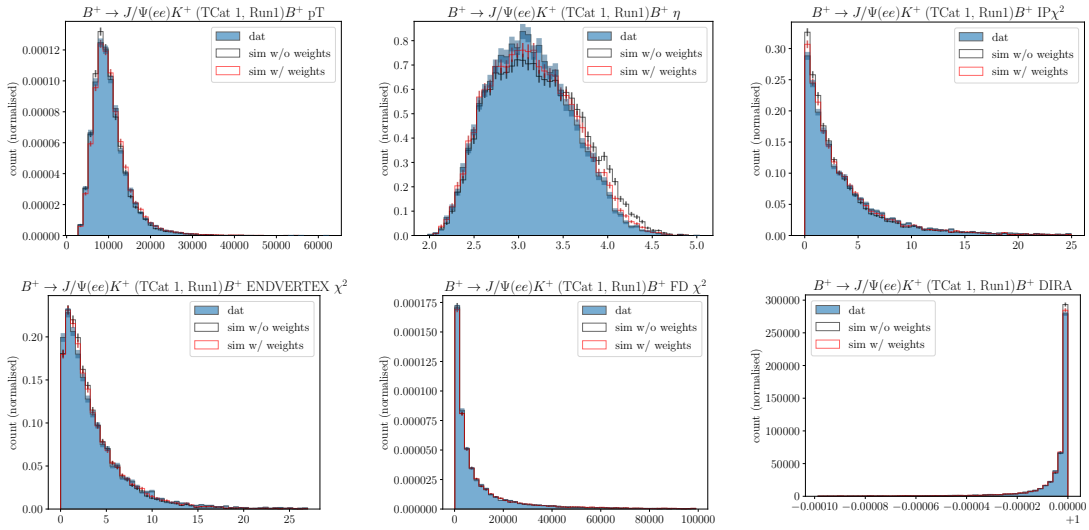

Figure A.2: The mis-modelling between the fully selected (including a cut on the constrained mass) simulated control mode samples and data is shown in the six input variables to `GBReweighter`. The data is shown in blue, while the simulated samples without weights shown in black, and the same simulated samples with the `FinalWeights` applied shown in red. The agreement is clearly strongly improved across all six variables. Run1 - hTOS.

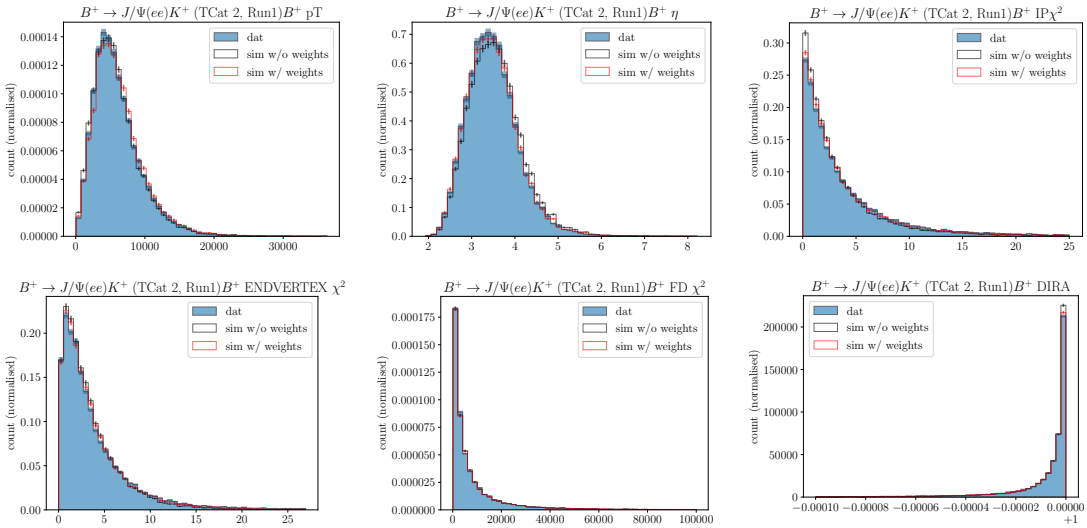

Figure A.3: The mis-modelling between the fully selected (including a cut on the constrained mass) simulated control mode samples and data is shown in the six input variables to `GBReweighter`. The data is shown in blue, while the simulated samples without weights shown in black, and the same simulated samples with the `FinalWeights` applied shown in red. The agreement is clearly strongly improved across all six variables. Run1 - TIS.

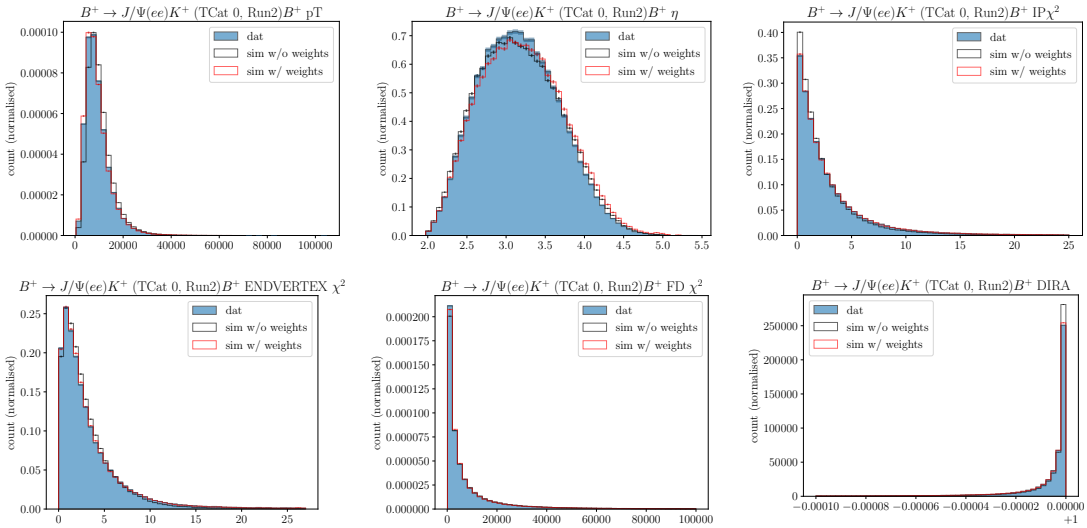

Figure A.4: The mis-modelling between the fully selected (including a cut on the constrained mass) simulated control mode samples and data is shown in the six input variables to `GBReweighter`. The data is shown in blue, while the simulated samples without weights shown in black, and the same simulated samples with the `FinalWeights` applied shown in red. The agreement is clearly strongly improved across all six variables. Run2 - eTOS.

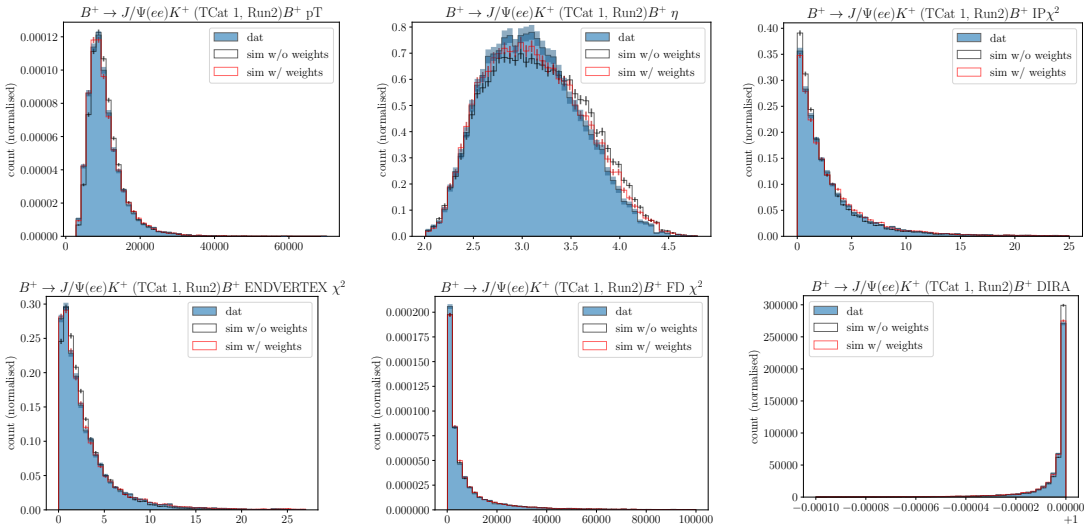

Figure A.5: The mis-modelling between the fully selected (including a cut on the constrained mass) simulated control mode samples and data is shown in the six input variables to `GBReweighter`. The data is shown in blue, while the simulated samples without weights shown in black, and the same simulated samples with the `FinalWeights` applied shown in red. The agreement is clearly strongly improved across all six variables. Run2 - hTOS.

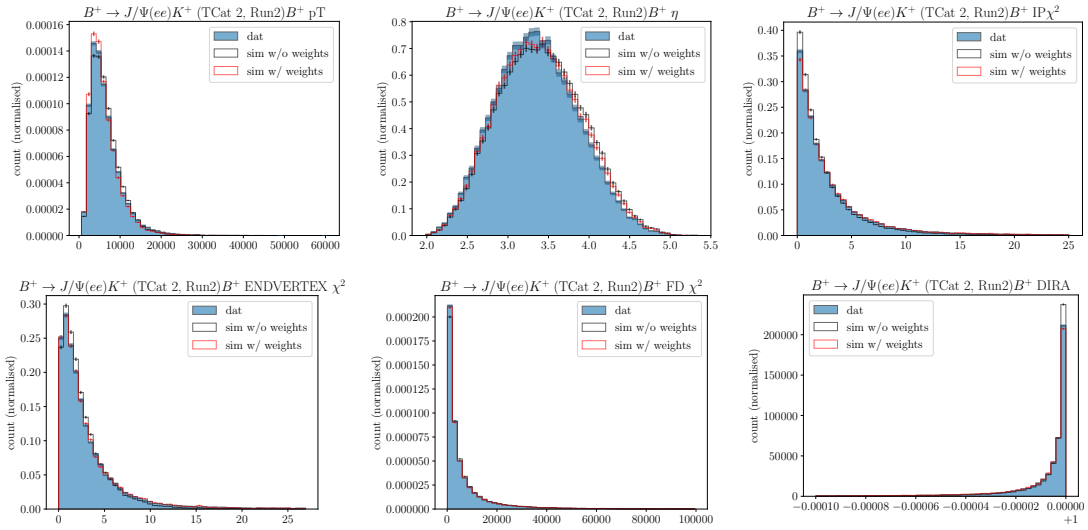

Figure A.6: The mis-modelling between the fully selected (including a cut on the constrained mass) simulated control mode samples and data is shown in the six input variables to `GBReweighter`. The data is shown in blue, while the simulated samples without weights shown in black, and the same simulated samples with the `FinalWeights` applied shown in red. The agreement is clearly strongly improved across all six variables. Run2 - TIS.

A.2 Mass Fits

A.2.1 Control Mode

The control mode simulated samples and experimental data are used to obtain the line shapes to fit the reconstructed mass. Each Bremsstrahlung and trigger category is fitted independently for each run. The simulated samples and data are fitted simultaneously, using the same model to describe the line shape, with the addition of two scaling terms Δ_μ and Δ_θ that accommodate the mis-match between simulation and data. The fits to the control mode are given in Fig. [A.7, A.12].

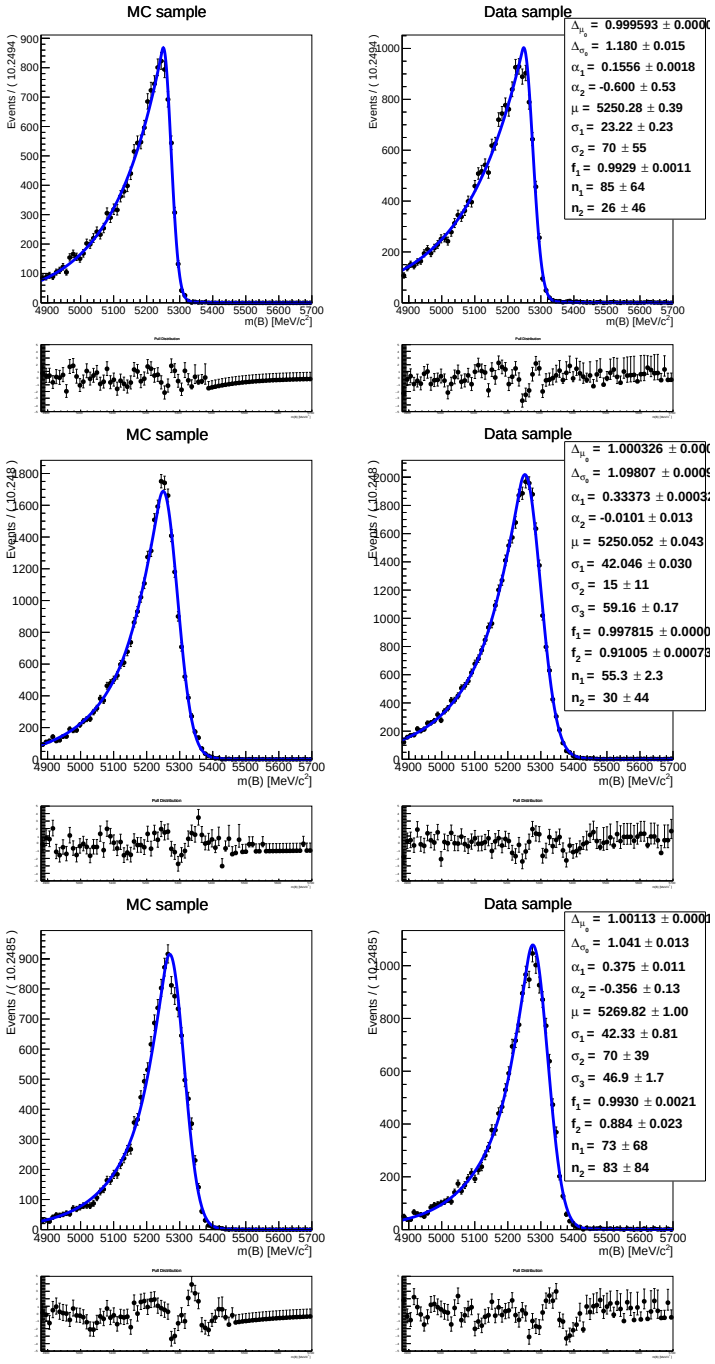

Figure A.7: **Trigger Category ETOS (Run 1)** Showing the simultaneous fits to the control mode simulation samples (left) and the mass constrained data (right), in three separate rows of bremsstrahlung recovery: 0γ (top); 1γ (middle); 2γ (bottom) for the eTOS trigger category in Run1. All parameters are shared per bremsstrahlung recovery category, as explained in Sec. 4.7, with two additional terms Δ_{μ_i}, and Δ_{σ_i} describing the multiplicative factor used for the central value (μ) and the widths (σ_i).

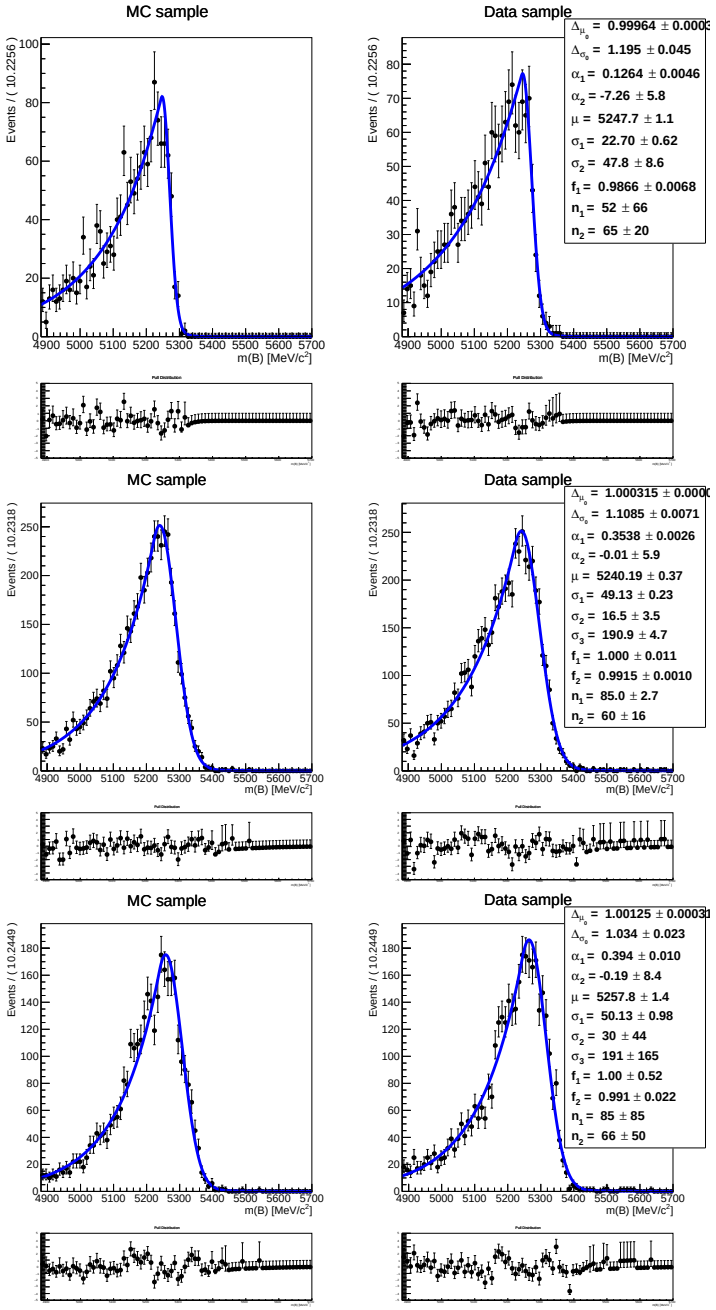

Figure A.8: **Trigger Category HTOS (Run 1)** Showing the simultaneous fits to the control mode simulation samples (left) and the mass constrained data (right), in three separate rows of bremsstrahlung recovery: 0γ (top); 1γ (middle); 2γ (bottom) for the hTOS trigger category in Run1. All parameters are shared per bremsstrahlung recovery category, as explained in Sec. 4.7, with two additional terms Δ_{μ_i}, and σ_{s_i} describing the multiplicative factor used for the central value (μ) and the widths (σ_i).

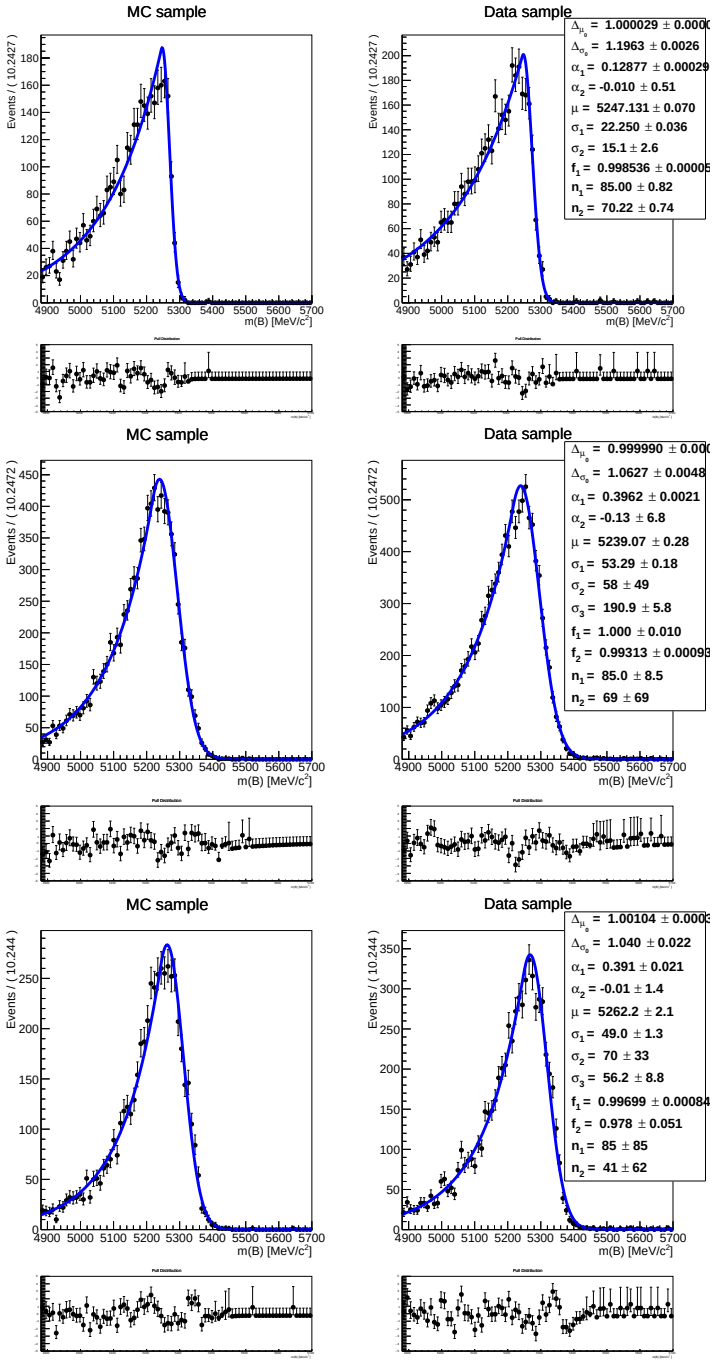

Figure A.9: **Trigger Category TIS (Run 1)** Showing the simultaneous fits to the control mode simulation samples (left) and the mass constrained data (right), in three separate rows of bremsstrahlung recovery: 0γ (top); 1γ (middle); 2γ (bottom) for the TIS trigger category in Run1. All parameters are shared per bremsstrahlung recovery category, as explained in Sec. 4.7, with two additional terms Δ_{μ_i}, and σ_{s_i} describing the multiplicative factor used for the central value (μ) and the widths (σ_i).

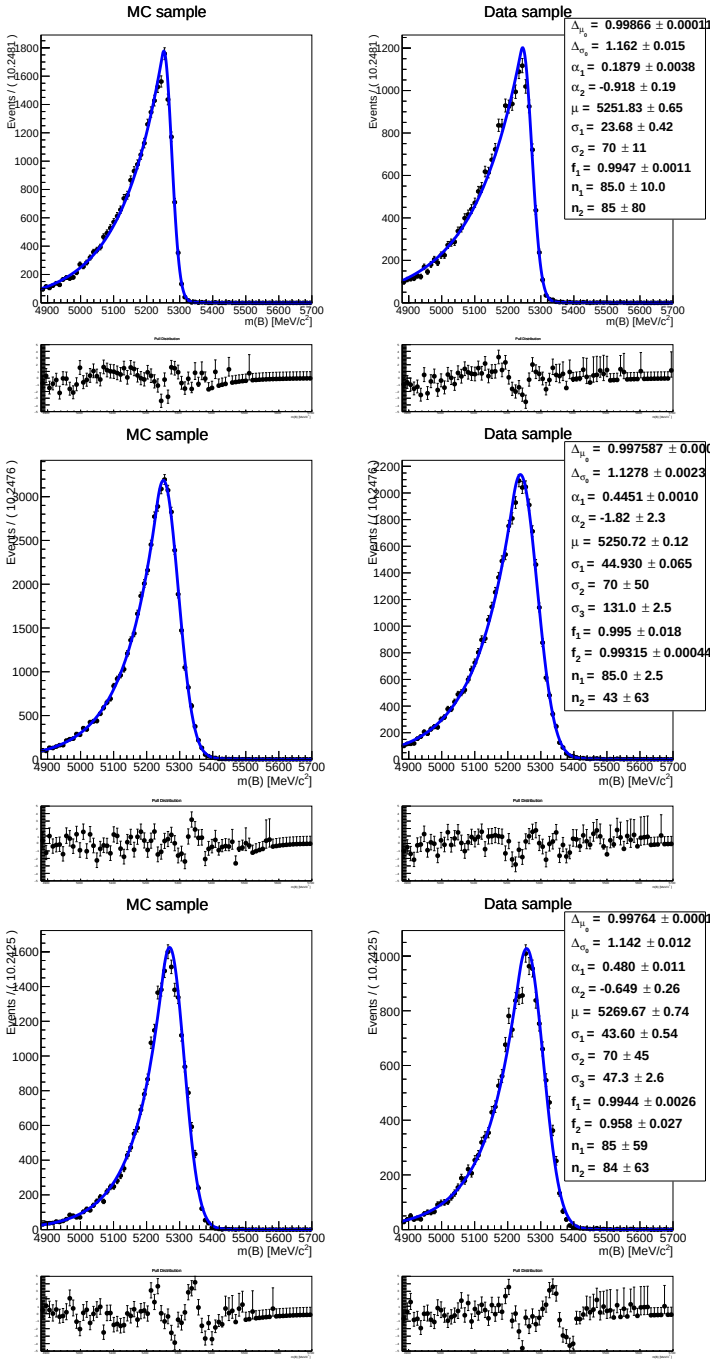

Figure A.10: **Trigger Category ETOS (Run 2)** Showing the simultaneous fits to the control mode simulation samples (left) and the mass constrained data (right), in three separate rows of bremsstrahlung recovery: 0γ (top); 1γ (middle); 2γ (bottom) for the eTOS trigger category in Run2. All parameters are shared per bremsstrahlung recovery category, as explained in Sec. 4.7, with two additional terms Δ_{μ_i}, and σ_{s_i} describing the multiplicative factor used for the central value (μ) and the widths (σ_i).

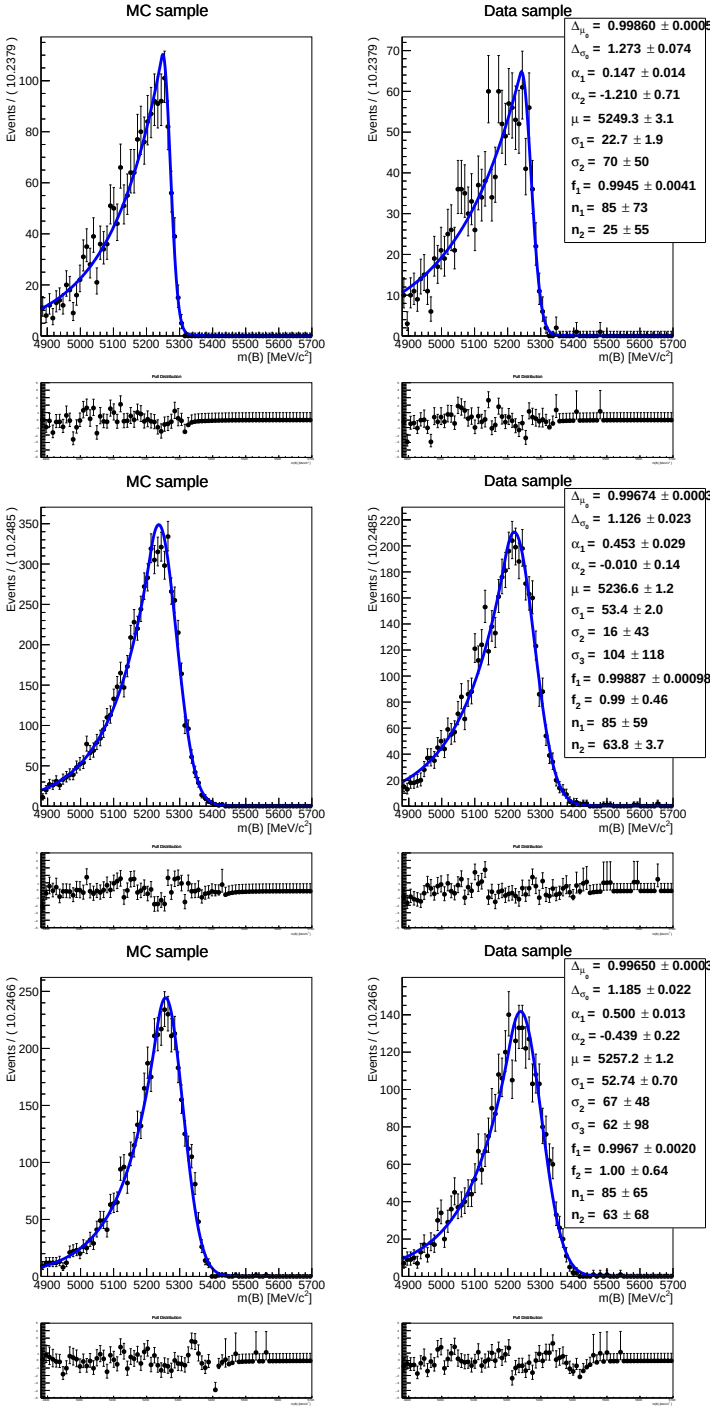

Figure A.11: **Trigger Category HTOS (Run 2)** Showing the simultaneous fits to the control mode simulation samples (left) and the mass constrained data (right), in three separate rows of bremsstrahlung recovery: 0γ (top); 1γ (middle); 2γ (bottom) for the hTOS trigger category in Run2. All parameters are shared per bremsstrahlung recovery category, as explained in Sec. 4.7, with two additional terms Δ_{μ_i}, and σ_{s_i} describing the multiplicative factor used for the central value (μ) and the widths (σ_i).

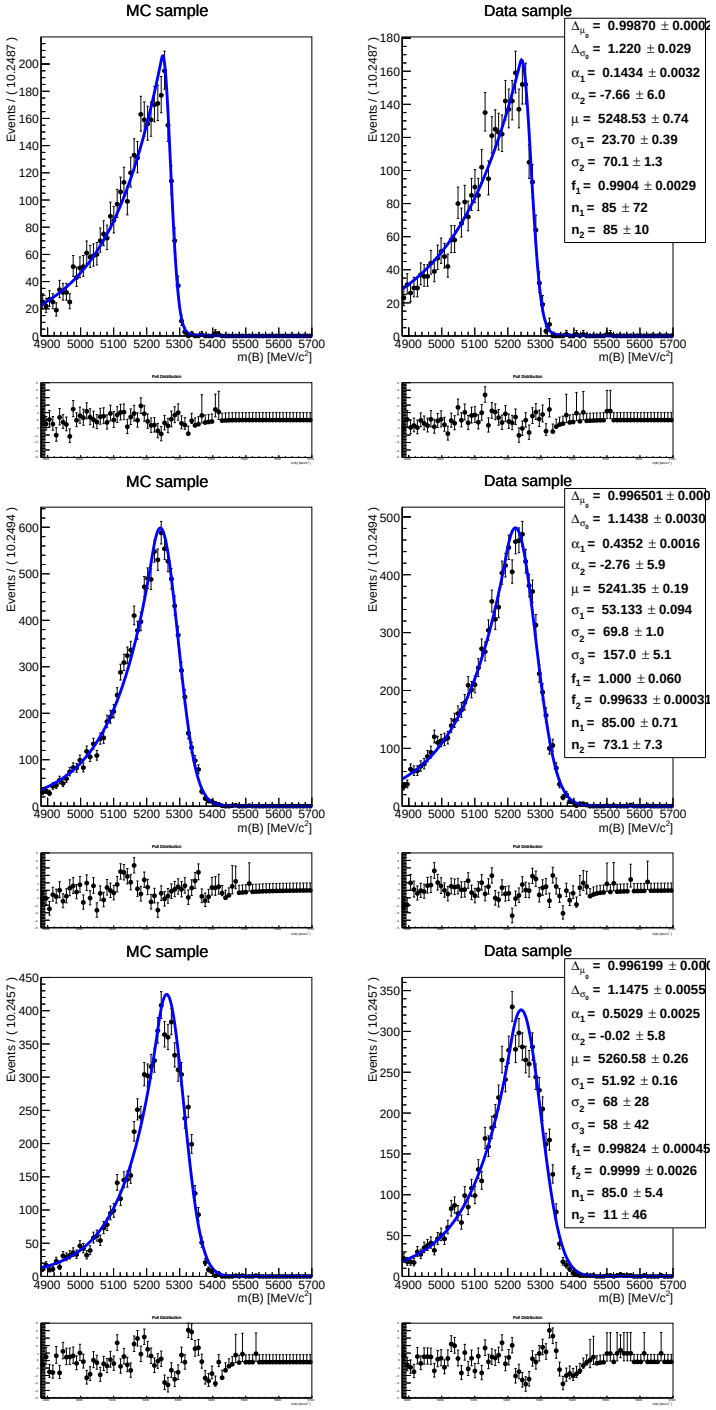

Figure A.12: **Trigger Category TIS (Run 2)** Showing the simultaneous fits to the control mode simulation samples (left) and the mass constrained data (right), in three separate rows of bremsstrahlung recovery: 0γ (top); 1γ (middle); 2γ (bottom) for the eTOS trigger category in Run2. All parameters are shared per bremsstrahlung recovery category, as explained in Sec. 4.7, with two additional terms Δ_{μ_i}, and σ_{s_i} describing the multiplicative factor used for the central value (μ) and the widths (σ_i).

A.2.1.1 Constrained Mass Study

Throughout this analysis it is useful to take advantage of a control mode region where it can be assumed there is a negligible combinatorial background contribution. This is used as a clean environment to obtain the simulation corrections in Sec. 4.5, and to obtain the transformation factors of the mean and widths of the signal mass fit shapes in Sec. 4.7, as well as various additional cross checks and studies. To obtain the clean background free control mode data sample the constrained mass is used with a selection placed on it as $5185 < m_{const}$ MeV when performing mass fits and cross checks and as $5185 < m_{const} < 5350$ MeV when obtaining the simulation corrections. To demonstrate that this sample is completely background free the control mode mass was fitted with the mass constraint selection applied, but with an additional combinatorial component floating. This is identical to the fitting procedure in Sec. 4.7 for the control mode, but with the partially-reconstructed component removed from the fit, giving a sense of how much combinatorial background was present in the fit. This study demonstrates that the background yield as determined by the fitting procedure is compatible with 0 for each trigger category and run period in the $m_{HOP} > 4900$ MeV region, with results shown in Fig. A.13.

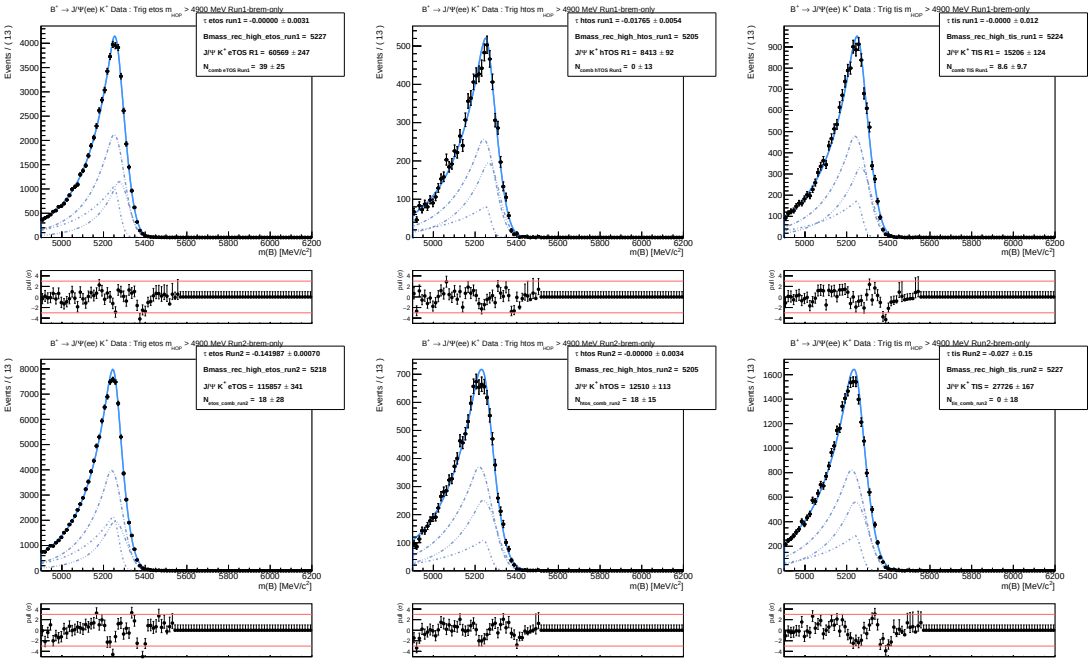

Figure A.13: Run1 (top) and Run2 (bottom) fits to the $m_{HOP} > 4900\,\text{MeV}$ region for $B^+ \to J/\Psi(e^+e^-)K^+$ fully selected and mass constrained ($m_{const} > 5185\,\text{MeV}$) data, split into trigger categories $eTOS$ (left), $hTOS$ (centre), TIS (right). The combinatorial background would be shown in red, although the yield is compatible with zero in all cases. Each individual bremsstrahlung category as a dashed blue line, and the total model as a solid blue line.

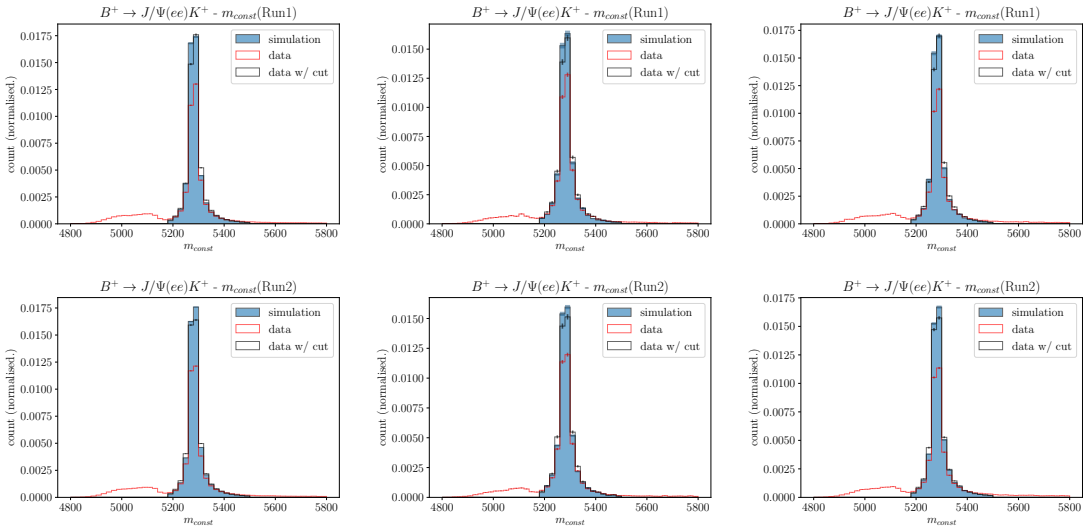

Figure A.14: Run1 (top) and Run2 (bottom) m_{const} distributions for control mode fully selected simulation and data, split into trigger categories $eTOS$ (left), $hTOS$ (centre), TIS (right). Simulation is shown in blue, data shown in red prior to any m_{const} selection, and in black the data with a selection of $m_{const} > 5185\,\text{MeV}$ placed.

A.2.1.2 Mass Constraint Distributions

The mass constraint distributions used to obtain the selections in Sec.A.2.1.1 are shown in this section, where the full m_{const} range in the control mode data is shown, broken into three trigger categories and the two run periods, compared against the simulated control mode samples. Both sets of samples have the m_{const} selection applied, and as can be seen in Fig. A.14 the data and simulation samples are in close agreement.

A.2.2 Rare Mode Simulation

The rare mode simulated samples are fitted to obtain the line-shape used in the mass fit to the data. These fits are done per trigger, brem, and run combination, resulting in 18 signal line-shapes. Figures [A.15 - A.20] detail each of these fits. The model in each case is the same as the model used in the control mode, without the additional transformation factor the describes the shift to the mean and width between simulation and data. In the rare mode fits to data the values are taken from the fits to the control mode. Finally Fig.A.21, and Fig.A.22 show the combined fit to each trigger category for Run1 and Run2 simulated sample. The combination of bremsstrahlung categories models the data well.

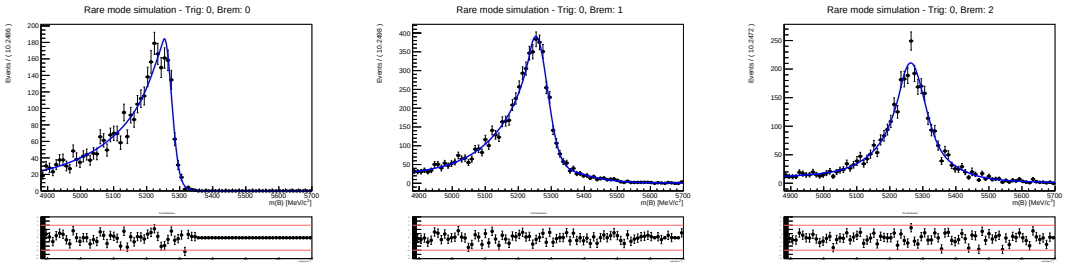

Figure A.15: Run 1 fits to the Kee *eTOS* fully selected simulated samples for bremsstrahlung categories $\gamma = 0, 1, 2$.

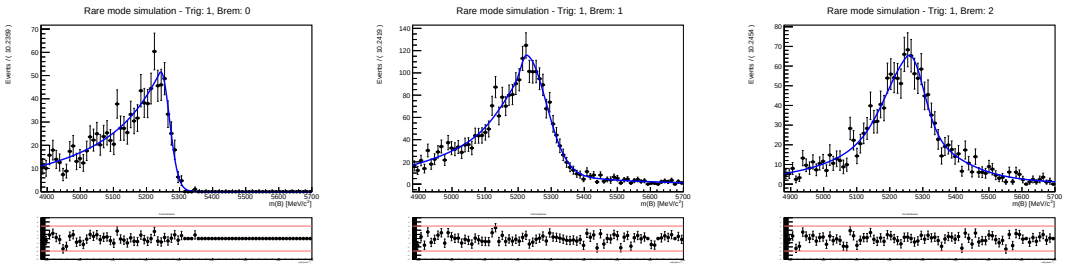

Figure A.16: Run 1 fits to the Kee *hTOS* fully selected simulated samples for bremsstrahlung categories $\gamma = 0, 1, 2$.

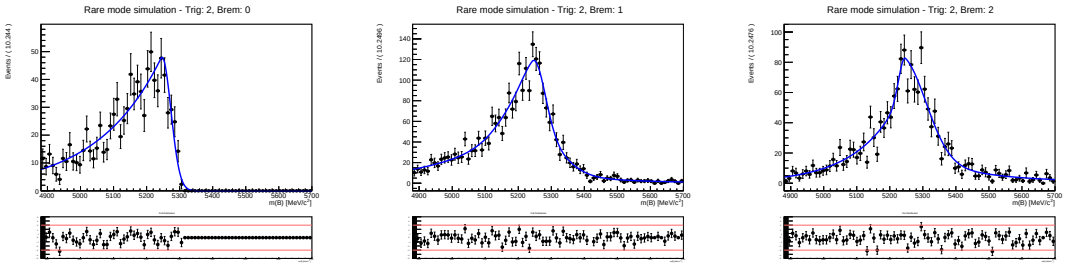

Figure A.17: Run 1 fits to the Kee TIS fully selected simulated samples for bremsstrahlung categories $\gamma = 0, 1, 2$.

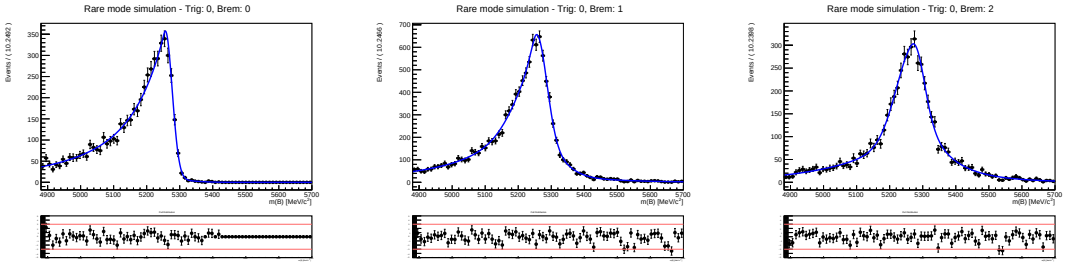

Figure A.18: Run 2 fits to the Kee $eTOS$ fully selected simulated samples for bremsstrahlung categories $\gamma = 0, 1, 2$.

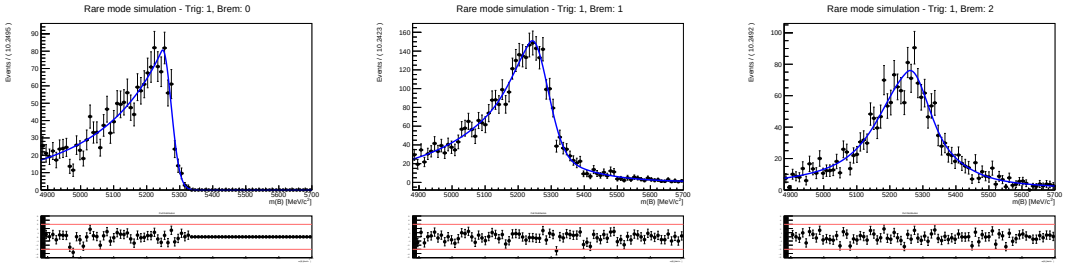

Figure A.19: Run 2 fits to the Kee $hTOS$ fully selected simulated samples for bremsstrahlung categories $\gamma = 0, 1, 2$.

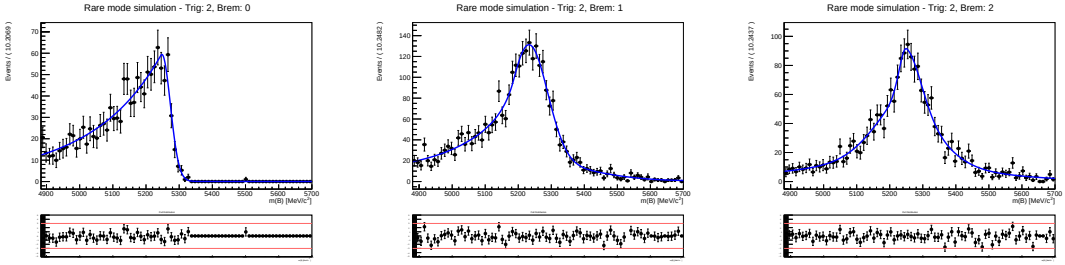

Figure A.20: Run 2 fits to the Kee TIS fully selected simulated samples for bremsstrahlung categories $\gamma = 0, 1, 2$.

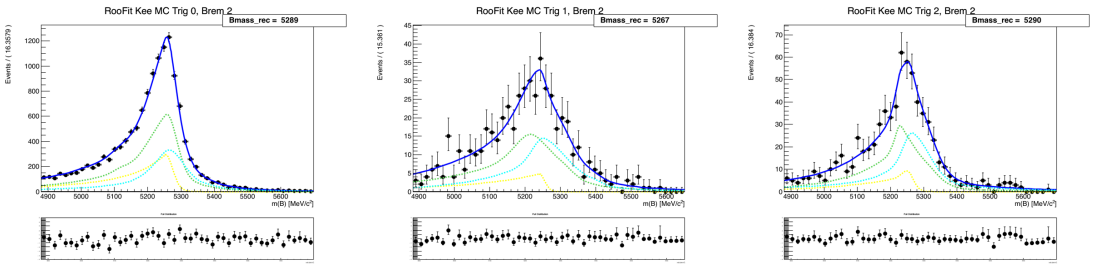

Figure A.21: Mass fits to the $B^+ \to K^+e^+e^-$ Run 2 simulated, fully selected, samples. Split by trigger category, eTOS (left), hTOS (centre) and TIS (left), show with the combination of Bremstrahlung categories forming the dark blue line shape to fit the data.

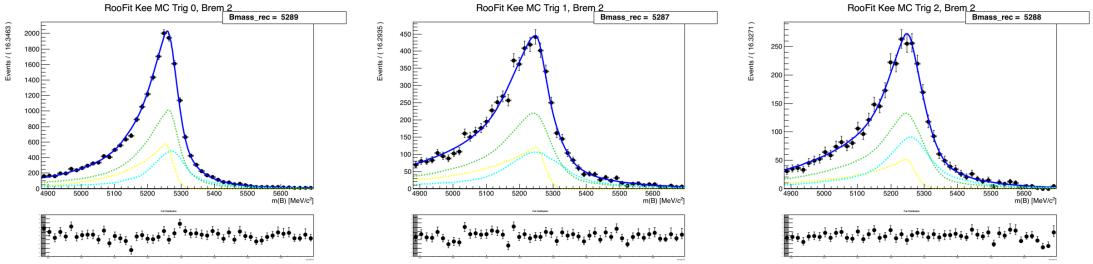

Figure A.22: Mass fits to the $B^+ \to K^+e^+e^-$ Run 2 simulated, fully selected, samples. Split by trigger category, eTOS (left), hTOS (centre) and TIS (left), show with the combination of Bremstrahlung categories forming the dark blue line shape to fit the data.

A.2.2.1 Mass fit pull studies

A pull study was undertaken to evaluate the performance of the mass fitting procedure and to evaluate bias. This consisted of generating toy datasets from the nominal fit model, and recording the pull value for each of the yields. A Pull distribution centred at 0, following a Gaussian distribution with width $\sigma = 1$ is considered good. The results of these studies for Run1 and Run2 are shown in Figs.[A.23, A.23].

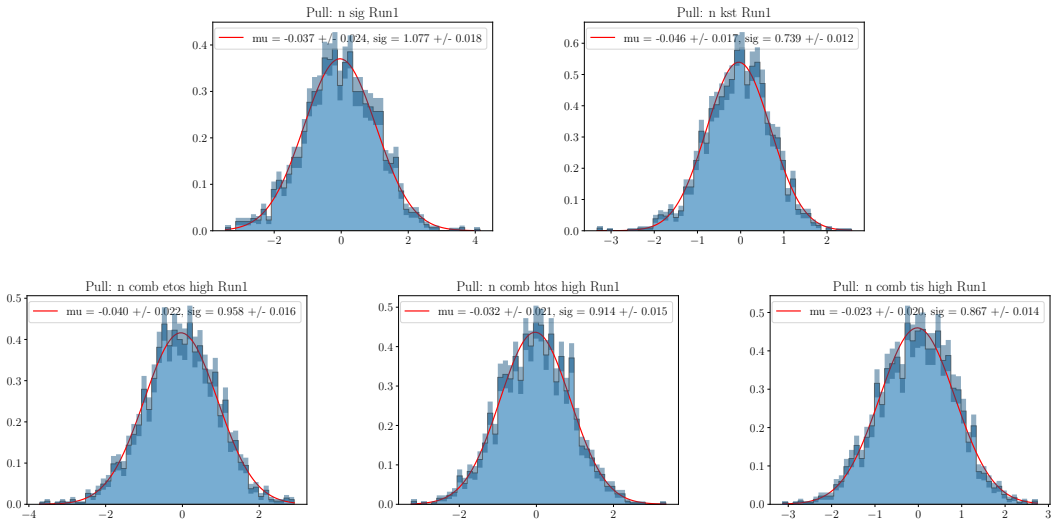

Figure A.23: Pull studies of the mass fit yields in Run1 $B^+ \to K^+e^+e^-$. Distributions show for the signal yield (top left), the partially reconstructed yield (top right), and the combinatorial yield for the combinatorial yield in eTOS (bottom left), hTOS (bottom centre) and TIS (bottom right). Each has a Gaussian fitted with results shown displayed.

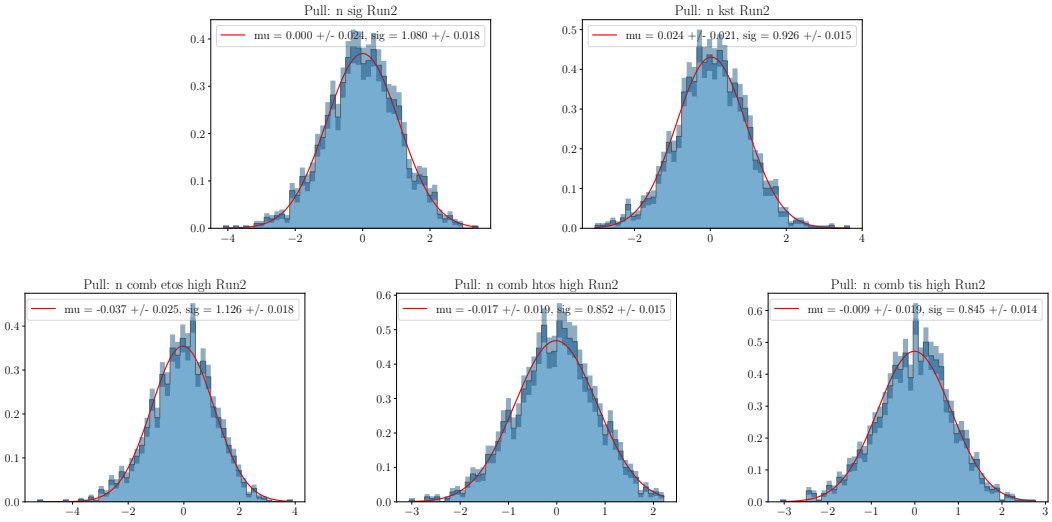

Figure A.24: Pull studies of the mass fit yields in Run1 $B^+ \to K^+e^+e^-$. Distributions show for the signal yield (top left), the partially reconstructed yield (top right), and the combinatorial yield for the combinatorial yield in eTOS (bottom left), hTOS (bottom centre) and TIS (bottom right). Each has a Gaussian fitted with results shown displayed.

A.3 Mass Line-Shape q^2 dependence

The independence of the mass fit line-shape on q^2 is important as the reconstructed mass distribution in the control mode is used to obtain corrections that are applied to the rare mode mass fit line shapes. The dependence of the mass shape on the q^2 region was evaluated in two further ways. Firstly looking at the mass distribution of $B^+ \to K^+e^+e^-$ simulated samples in slices of q^2, shown in Figs.[A.25 - A.26], which demonstrates no q^2 dependence in the 5 GeV wide rare mode window. Secondly by comparing the core of the mass distribution in $B^+ \to K^+e^+e^-$ and $B^+ \to J/\Psi(e^+e^-)K^+$ simulated samples, shown in Fig.[A.27-A.28], which demonstrates the consistency of the core of the mass distribution over a range of approximately 9 GeV.

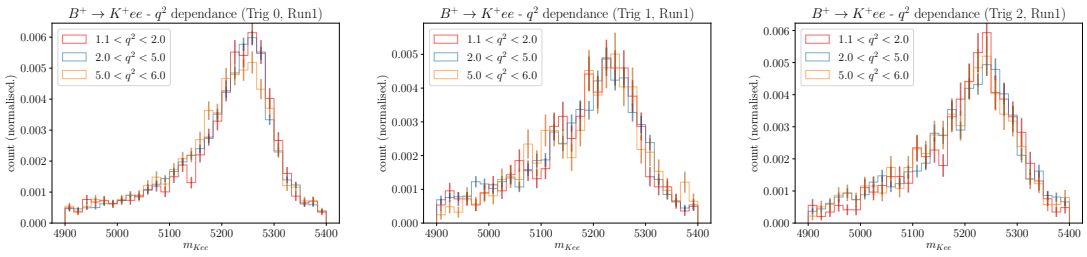

Figure A.25: The $B^+ \to K^+e^+e^-$ simulated samples mass distribution, split by trigger category, shown in three slices of q^2. (Run1)

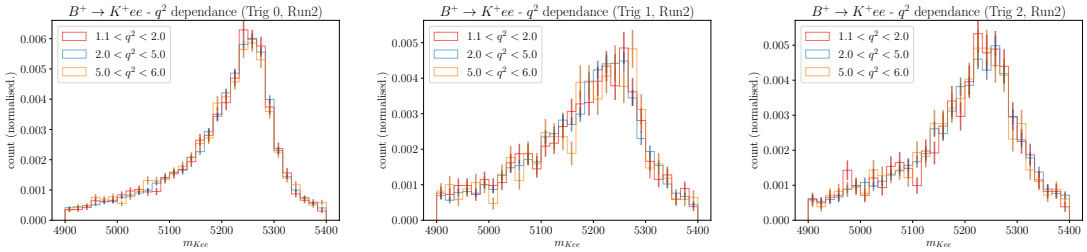

Figure A.26: The $B^+ \to K^+e^+e^-$ simulated samples mass distribution, split by trigger category, shown in three slices of q^2. (Run2)

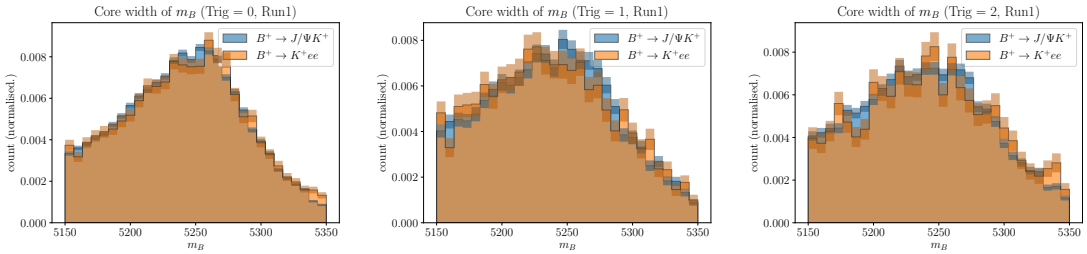

Figure A.27: The $B^+ \to K^+e^+e^-$ and $B^+ \to J/\Psi(e^+e^-)K^+$ simulated samples core mass distribution, split by trigger category, showing the independence of q^2. (Run1)

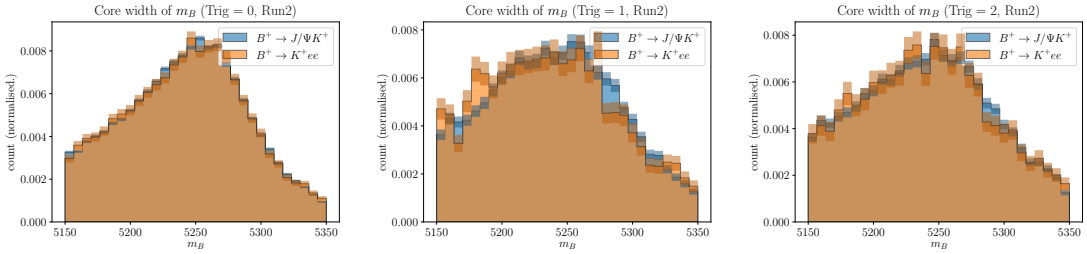

Figure A.28: The $B^+ \to K^+e^+e^-$ and $B^+ \to J/\Psi(e^+e^-)K^+$ simulated samples core mass distribution, split by trigger category, showing the independence of q^2. (Run2)

A.4 Background subtraction Residuals and Spread verification

When generating toy background samples to subtract in the Feldman Cousins method in Sec. 4.10.3 it is important to verify that the spread of each component follows a normal distribution with the expected spread based on the uncertainty in the fit to the mass distribution, and that when subtracted from the background template gives a central value of 0 events. Figure Fig. A.29 shows that when the background subtraction is performed the residual number of events shows no bias away from a count of 0 events.

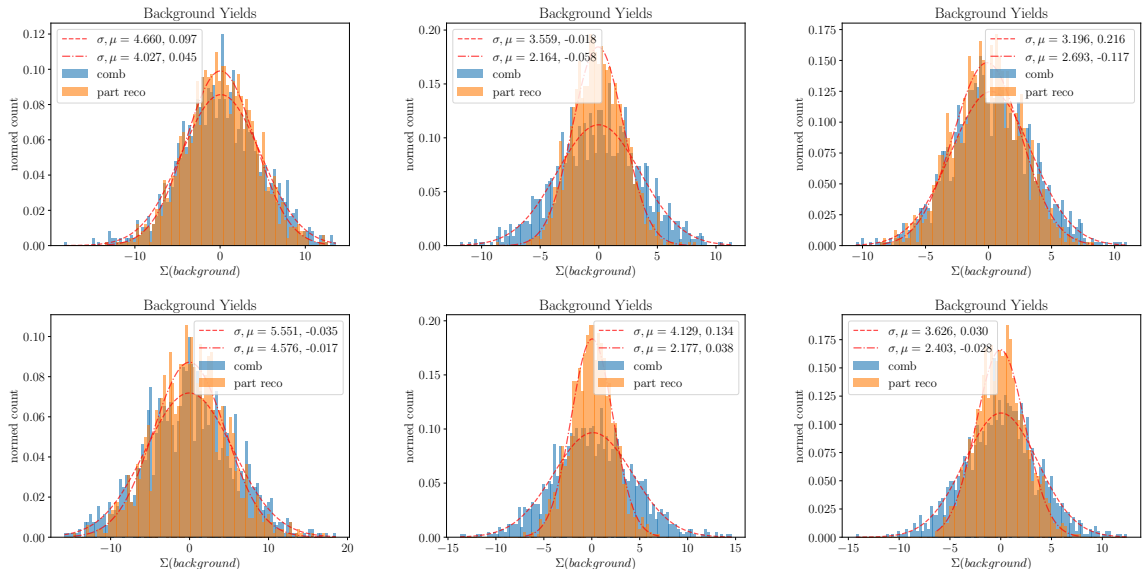

Figure A.29: Spread of residual yields of generated backgrounds for Run1 (top) and Run2 (bottom) for eTOS (left) hTOS (centre) and TIS (right). These background have been generated from the multivariate Gaussian distribution that incorporates the correlations taken from the rare mode fit, then have the nominal background template subtracted. Each plot shows the spread of both combinatorial and partially reconstructed backgrounds, and is fitted with a Gaussian. All data are well described, and centred at 0.

		Signal			Comb			Part-Reco		
		eTOS	hTOS	TIS	eTOS	hTOS	TIS	eTOS	hTOS	TIS
Signal	eTOS	1.000	0.000	0.000	0.078	0.000	0.000	-0.378	0.000	0.000
	hTOS		1.000	0.000	0.000	-0.384	0.000	0.000	-0.055	0.000
	TIS			1.000	0.000	0.000	0.089	0.000	0.000	-0.417
Comb	eTOS				1.000	0.000	0.000	-0.749	0.000	0.000
	hTOS					1.000	0.000	0.000	-0.036	0.000
	TIS						1.000	0.000	0.000	-0.727
Part-Reco	eTOS							1.000	0.000	0.000
	hTOS								1.000	0.000
	TIS									1.000

Table A.1: Run1 Control mode correlations

		Signal			Comb			Part-Reco		
		eTOS	hTOS	TIS	eTOS	hTOS	TIS	eTOS	hTOS	TIS
Signal	eTOS	1.000	0.000	0.000	-0.263	0.000	0.000	-0.075	0.000	0.000
	hTOS		1.000	0.000	0.000	-0.158	0.000	0.000	-0.323	0.000
	TIS			1.000	0.000	0.000	-0.248	0.000	0.000	-0.117
Comb	eTOS				1.000	0.000	0.000	-0.689	0.000	0.000
	hTOS					1.000	0.000	0.000	-0.063	0.000
	TIS						1.000	0.000	0.000	-0.714
Part-Reco	eTOS							1.000	0.000	0.000
	hTOS								1.000	0.000
	TIS									1.000

Table A.2: Run2 Control mode correlations

A.5 Mass fit correlation tables

In this section the correlation matrices for the control mode and rare mode mass fits are given. These matrices show the degree to which different variables are linked in the mass fit.

A.5.1 Control Mode Correlation Factors

Correlation factors for control mode signal and background yields are shown in Tab. A.1 and Tab. A.2, no correlations are greater than 0.75.

		Signal	Comb			Part-Reco
		combined	eTOS	hTOS	TIS	combined
Signal	combined	1.000	-0.528	-0.407	-0.414	0.172
Comb	eTOS	-0.528	1.000	0.364	0.401	-0.552
	hTOS	-0.407	0.364	1.000	0.293	-0.384
	TIS	-0.414	0.401	0.293	1.000	-0.456
Part-Reco	combined	0.172	-0.552	-0.384	-0.456	1.000

Table A.3: Run1 Rare mode correlations

		Signal	Comb			Part-Reco
		combined	eTOS	hTOS	TIS	combined
Signal	combined	1.000	-0.542	-0.418	-0.379	0.053
Comb	eTOS	-0.542	1.000	0.305	0.286	-0.369
	hTOS	-0.418	0.305	1.000	0.213	-0.252
	TIS	-0.379	0.286	0.213	1.000	-0.256
Part-Reco	combined	0.053	-0.369	-0.252	-0.256	1.000

Table A.4: Run2 Rare mode correlations

A.5.2 Rare Mode Correlation Factors

Correlations between rare mode signal and backgrounds yields are can be compared to those in the control mode in Sec.A.5.1, however despite a relative strong correlation, none are greater than 0.60, as seen in Tab. A.3 and Tab. A.3.

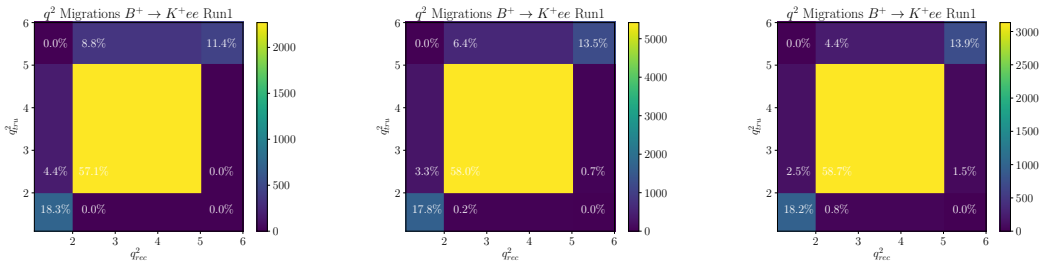

Figure A.30: Density plots showing the distribution patterns for each Bremsstrahlung category ($\gamma = 0, 1, 2$ left to right)) for Run1 rare mode simulated samples across true q^2 and reconstructed q^2 Only minimal differences are seen.

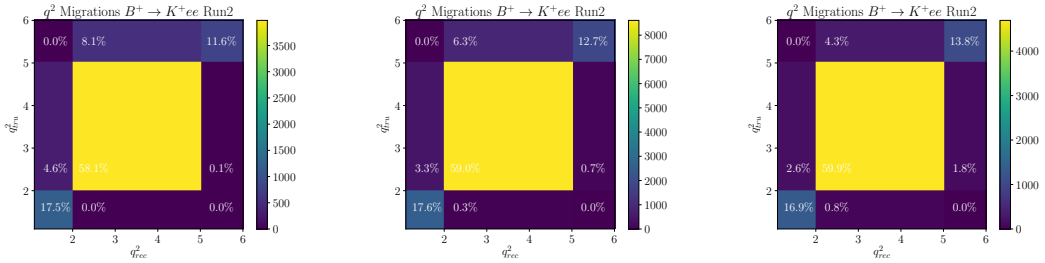

Figure A.31: Density plots showing the distribution patterns for each Bremsstrahlung category ($\gamma = 0, 1, 2$ left to right)) for Run2 rare mode simulated samples across true q^2 and reconstructed q^2 Only minimal differences are seen.

A.6 Unfolding model per Bremsstrahlung Category

The nominal unfolding method in Sec. 4.8 is not dependent on the Bremsstrahlung category, or Run period. As can be seen Fig. A.30 and Fig. A.31 the differences between the categories and run periods is small.

A.7 ProbNN for cascading semi-leptonic Backgrounds

This section is preserved to show the original method used to remove cascading semi-leptonic backgrounds. A neural network was used to select semi-leptonic events from rare mode signal events, as this would allow a simple model of the entire angular distribution to be used. This would have been beneficial as preserving the edge of the angular distribution close to $\cos\theta_\ell = 1$ *would have increased precision. The complexities of the semi-leptonic backgrounds made them hard to model effectively, and the neural network was not deemed sufficiently effective to use in the final analysis.*

The cascading semi-leptonic decays provide an additional source of background, they are primarily caused by decays with an intermediate state including a D^0 that decays to a final state including either neutrinos or pions, which can be missed or mis-identified. Typically these events are removed using a veto, a combination of $m(K^+l^-) > 1885$ MeV requiring the events has to have a mass greater than the $m(D^0)$, and a narrow window around the $m(D^0)$ such that $m^{track}_{e\to\pi}(K^+e^-) \notin m(D^0) \pm 40$ MeV. In the R_K analysis this is modelled as an efficiency, and in the angular analysis of $B^+ \to K^+\mu^+\mu^-$ a narrow selection can be made on $m(K^+l-)$ [5]. The reasons that neither of these methods are applicable to this analysis and how the cascading semi-leptonic backgrounds were removed are detailed below.

The veto $m_{K^+\ell^-} < m_{D^0}$ detailed in section 4.6.2 causes a very sharp drop in the angular distribution at $cos\theta_l > 0.66$, so the angular distribution was initially fitted up to $cos\theta_l < 0.6$ to avoid the complexities in modelling the drop off, however this resulted in an approximate degeneracy between the observables A_{FB} and F_H giving rise to large uncertainties in these parameters on the order of several hundred times the expected values, without a clear global solution. It is not possible to circumvent these backgrounds as was done in [5] due to the very broad resolution of electrons compared to muons.

Therefore, a more focused selection against the semi-leptonic backgrounds, that replaces the cut on $m_{K^+\ell^-}$, was constructed that would allow an angular model to be fit to the full range $-1.0 < cos\theta_l < 1.0$. This does not remove the need for an asymmetric description of the angular distribution but provides enough of an additional constraint to bring the uncertainties under control. The impact of the semi-leptonic backgrounds is tightly constrained in the $0.6 < cos\theta_l < 0.97$ region, enabling a focused selection in this range. The additional focused selections consist of a dedicated neural network classifier trained to select signal- from semi-leptonic decays ($\text{ProbNN}_{\text{cascades}}$), a cut on $m_{l\to\pi}(K^+l^-) \notin m(D^0) \pm 40$ MeV, and tightening of the combinatorial classifier.

The $\text{ProbNN}_{\text{cascades}}$ classifier is the same architecture as the network used for the $\text{ProbNN}_{\text{shells}}$ classifier but with 10 inputs rather than 8, and trained using the same standards to avoid over-training. The input variables to the classifier are the B^+_{DIRA}, $B^+_{\text{IP}\chi^2}$, $B^+_{\text{FD}\chi^2}$, $B^+_{\text{EV}\chi^2}$, e^{max}_{pT}, e^{min}_{pT} P^{LOF}, P^{LOF}_T, $\theta_{\ell\ell}$, $\theta_{(\ell\ell)K}$. Where e^{max}_{pT} and e^{min}_{pT} are in minimum and maximum electron pT, P^{LOF} and P^{LOF}_T, are the momentum in the line of, and transverse to the line of, flight of the B, and $\theta_{\ell\ell}$ and $\theta_{(\ell\ell)K}$ are the opening angle of the di-lepton system and the di-lepton system with the Kaon.

The datasets used to train the $ProbNN_{\text{cascades}}$ classifier are $B^+ \to K^+e^+e^-$simulated samples as signal, and $B^+ \to D(\to K^+e\nu)e\nu$ simulated samples as background. The relative proportions and separation into test and train samples are detailed in Table.A.6. The input training variables and the separation between signal and background can be seen in Fig.A.32. The neural network output from the best model is shown in Fig.A.33 where the separation between signal and background is clearly visible, the validation output is overlaid and can be seen to be in good agreement indicating no over-training bias.

While $B^+ \to D(\to K^+e\nu)e\nu$ is not the only cascading semi-leptonic background in the $B^+ \to K^+e^+e^-$data sample, it is a dominant contribution, and therefore provides a good approximation of the background properties. Other methods were investigated: using a low mass sample from the data $m_B < 4900\,\text{MeV}$; synthetically boosting the dataset

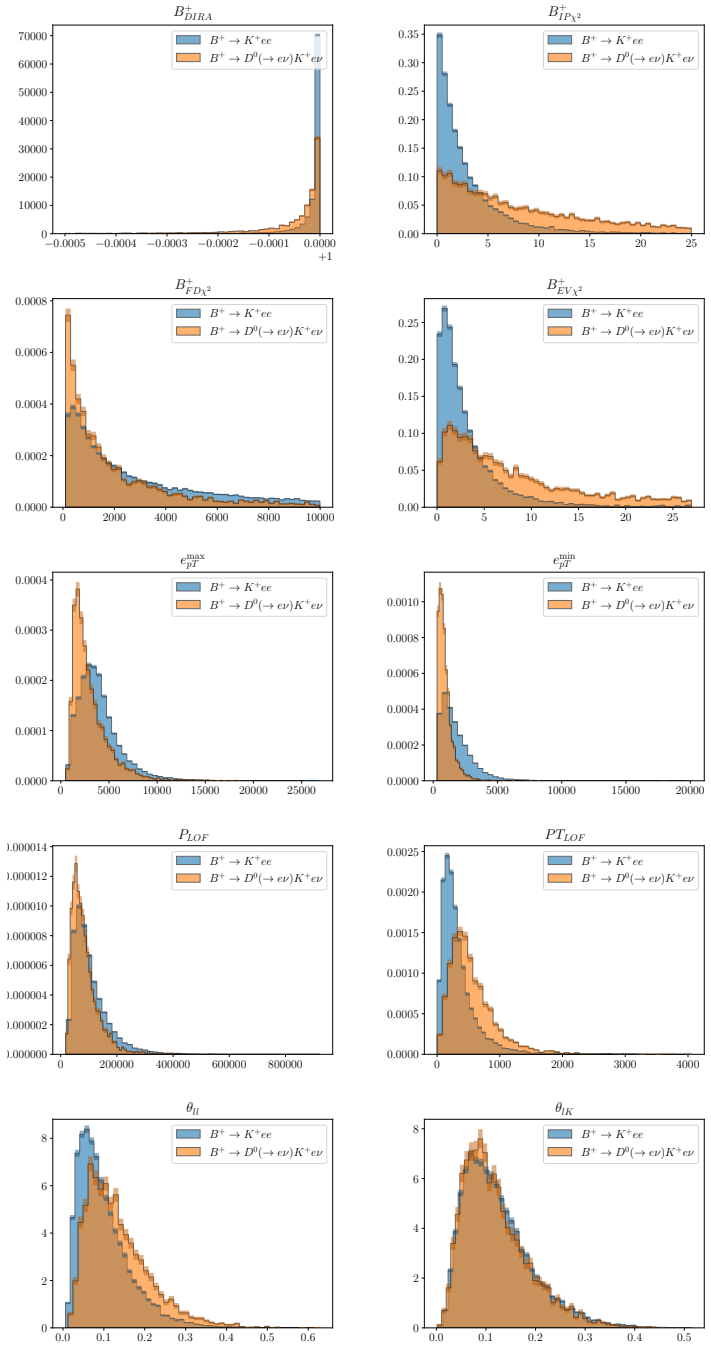

Figure A.32: The ten input variables to the $ProbNN_{\text{cascades}}$ classifier (B^+_{DIRA}, $B^+_{\text{IP}\chi^2}$, $B^+_{\text{FD}\chi^2}$, $B^+_{\text{EV}\chi^2}$, $e^{\max}_{pT}$, $e^{\min}_{pT}$ P^{LOF}, P^{LOF}_T, $\theta_{\ell\ell}$, $\theta_{(\ell\ell)K}$), showing the discrimination between $B^+ \to K^+e^+e^-$ and $B^+ \to D^0(\to K^+e\nu)e\nu$ simulated samples. To aid visualisation the $B^+ \to K^+e^+e^-$ MC sample is limited to the same size as the $B^+ \to D^0(\to K^+e\nu)e\nu$ sample.

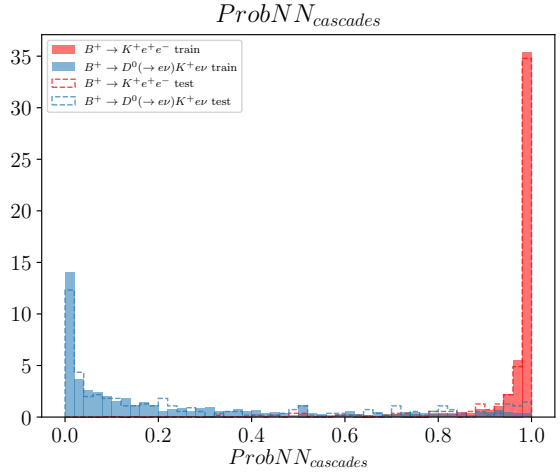

Figure A.33: $ProbNN_{cascades}$ output showing the separation between signal and background, for the test and train samples. The $B^+ \to K^+e^+e^-$ sample is three times larger than the $B^+ \to D^0(\to K^+e\nu)e\nu$ sample, and the test:train ratio is 20:80.

Run	decay	type	range	test:train	#
Run1 + Run2	$B \to D(\to K^+e\nu)e\nu$	MC	$4900 < m_B < 6200$ MeV	20:80	2770
Run2	$B^+ \to K^+e^+e^-$	MC	$4900 < m_B < 6200$ MeV	20:80	8310

Table A.5: Statistics used in the training of the $ProbNN_{cascades}$ classifier. Table shows the number of events from the signal sample ($B^+ \to K^+e^+e^-$simulation) and background sample ($B^+ \to D(\to K^+e\nu)e\nu$ simulated samples) split by run period and the test / train split. The entries are summarised as ratios and fractions at the end of each row.

using the SMOTE algorithm [81]; and using the $B \to Ke\mu$ data sample. However, none of these methods facilitated a large enough dataset to produce robust performance in real application tests.

The performance of the classifier is good, reaching a validation accuracy of 90.4% on a hold out set not seen in training representing 20% of the original dataset, a significant improvement on the `GBClassifier` baseline. The training and testing statistics are summarised in table A.5.

The performance of the $\text{ProbNN}_{\text{cascades}}$ classifier is studied to check for bias and over-training in MC for the semi-leptonic backgrounds, and the signal MC not used in training or validation. The $\text{ProbNN}_{\text{cascades}}$ distributions in each reflect the performance desired and a conservative working point was chosen to reject the most background like events without rejecting too large a fraction of mis-identified true signal candidates. To bolster this selection a tighter $\text{ProbNN}_{\text{combinatorial}}$ is placed as well which reduces the expected

semi-leptonic background yield to consistent with zero events. The impact of each of these selections on both the signal mode Monte Carlo and semi-leptonic experimental data can be seen in the figures A.35 and A.36 respectively. This motivates a focused selection in the $0.6 < \cos\theta_l < 0.97$ region using both the tighter ProbNN > 0.995 selection and the additional use of the $\text{ProbNN}_{\text{cascades}} > 0.993$ selection. The focused cuts in the $0.6 < \cos\theta_l < 0.97$ region result in expected yields for each of the key semi-leptonic backgrounds detailed in section 4.6 are very small to vanishing, as can be seen in table A.6. The full procedure to obtain these values is composed of the following steps: For the $B^+ \to D^0(\to K^+ e\nu)e\nu$ sample there are enough events to calculate the expected number per run in a similar way to Eq. 4.4, with the extra efficiency factor ϵ_{veto} to describe the selection using the veto. For $B^+ \to D^0(\to K^+\pi)e\nu$ and $B^+ \to D^0(\to K^+ e\nu)\pi$ decays there were not enough samples to obtain the expected number of events in the same way, so a two stage calculation was performed First, an expected number of events passing the pre-selection and the combinatorial ProbNN selection is calculated as:

$$< N_{!\text{veto}} >= 2 \times f_{B^+} \times \sigma_{b\bar{b}} \times L \times BF \times \epsilon_{\text{selection}} S_i, \tag{A.1}$$

where S_i is the scale factor weight for a single event in the MC when applied to the data. Then, on the pre-selected samples, with no combinatorial selection, an efficiency of the veto selection is obtained, ϵ_{veto}.

Finally, the expected number is calculated as:

$$< N_{\text{Run}} >=< N_{!\text{veto}} > \times \epsilon_{\text{veto}}. \tag{A.2}$$

Clearly, this method provides a worst case estimate, as it assumes that the veto efficiency and the combinatorial selection are completely facorisable, however it is known that this is not the case. However, given that the expected yields are small to vanishing this is an acceptable over-estimation, and these backgrounds are not considered to have an impact.

As the classifier was trained using the cascading semi-leptonic simulated samples, an additional test using the very low mass side band in the $B^+ \to K^+e^+e^-$data and the $B \to Ke\mu$ data was performed. The results are summarised in Table.A.7

Decay	BF	S_1	S_2	$< N_{\text{Run1}} >$	$< N_{\text{Run2}} >$
$B^+ \to D^0(\to K^+ e\nu)e\nu$	7.7×10^{-4}	1.485	1.355	0.817 ± 0.246	0.745 ± 0.225
$B^+ \to D^0(\to K^+\pi)e\nu$	8.6×10^{-4}	0.244	0.223	0.012 ± 0.012	0.011 ± 0.011
$B^+ \to D^0(\to K^+ e\nu)\pi$	16.5×10^{-5}	3.472	3.168	1.736 ± 0.549	1.584 ± 0.501

Table A.6: Branching fractions and expected yields from selection efficiencies for the key cascading semi-leptonic backgrounds. An efficiency of selection is found using simulation samples over the full $4900 < m_B < 6500\,\text{MeV}$ range then an expected yield $< N >$ is calculated.

Decay	Run	Region	# remaining
$B^+ \to K^+e^+e^-$	Run1 + Run2	$4800 < m_B < 4900$, $0.6 < \cos\theta_\ell < 0.97$	2 ± 2
$B \to Ke\mu$	Run1	$4900 < m_B < 6200$, $0.6 < \cos\theta_\ell < 0.97$	< 3 @ 95%

Table A.7: Statistics corresponding to the additional cross check using $B^+ \to K^+e^+e^-$data ($m_B < 4900\,\text{MeV}$) and $B \to Ke\mu$ data to check the performance of the veto using the $ProbNN_{cascades}$ classifier.

The set of focused selections against the semi-leptonic backgrounds mean that the angular fit can be applied across the full range $-1. < cos\theta_l < 1.$, where the step introduced in the $0.6 < cos\theta_l < 0.97$ region can be parameterised as an efficiency correction, and the degeneracy broken, bringing uncertainties under control. While the fraction of signal events recovered using this more sophisticated veto is still suppressed, it can be seen in Sec. 4.10 using toy studies that the degeneracy is broken and angular fit uncertainties brought under control. The impact of the chosen region in $cos\theta_l$ can be seen in the leftmost plot of Fig. A.34 where events in the 100 MeV mass window $4800 < m_B < 4900$ MeV in the combined rare mode data for Run1 and Run2 show the step wise application of selections. Firstly shown in blue the base pre-selection has a clear semi-leptonic peak for higher values of $cos\theta_l$, which becomes more pronounced with the application of the ProbNN selection. This peak is almost purely due to semi-leptonic backgrounds, and can be entirely removed in this data sample with the focused selections. This study, combined with the Monte Carlo study in Table A.6, implies that the semi-leptonic backgrounds are understood and are negligible.

The applicability of this selection to the rare mode is ensured by using this $ProbNN_{cascade}$ selection in the angular fits to the $B^+ \to J/\Psi K^+$ mode, where the veto is superfluous to requirement. This study verified that the selection is well modelled in simulation samples and data, full details can be found in Sec. 4.10 detailing the angular fitting procedure.

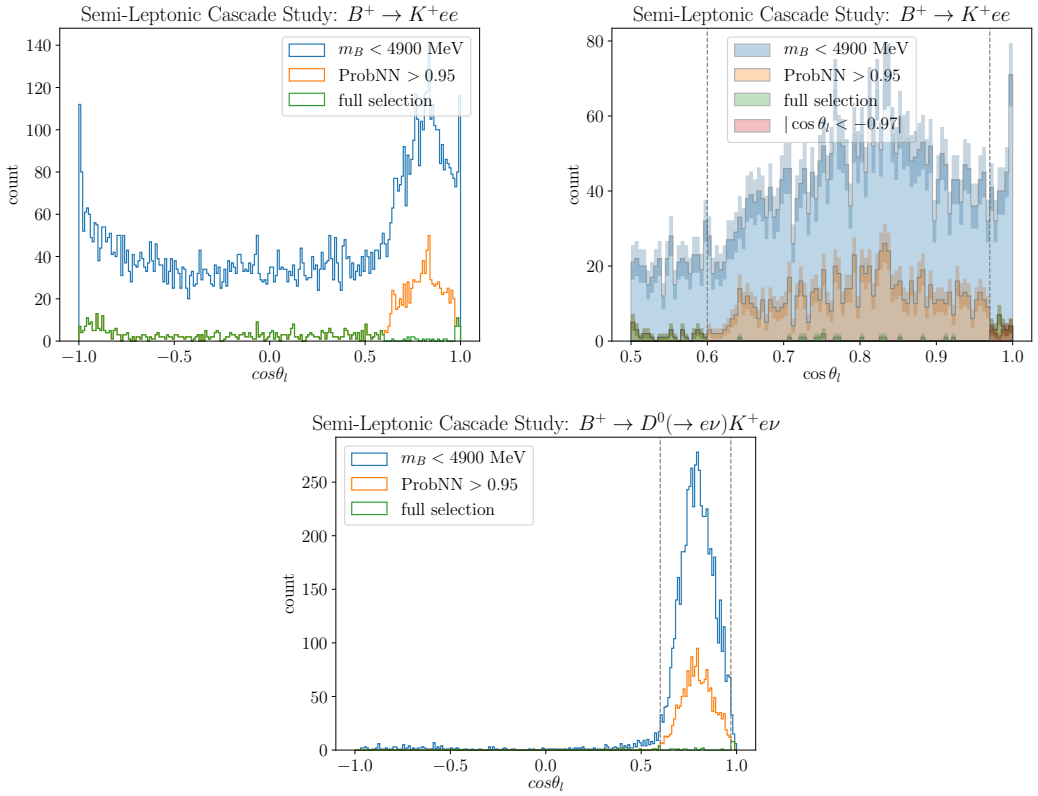

Figure A.34: Looking at the angular distribution for the combined Run1 and Run2 rare mode dataset in the lower mass side band ($4800 < m_B < 4900$) (top left) the veto on the semi-leptonic decays shows a clear reduction. The same data is presented in a zoomed in region (top right) with the lower region $|\cos\theta_\ell| < -0.97$ reflected onto the positive region. These two sections in green and red above the dashed line look to be in good agreement. Similarly in the $B^+ \to D^0(\to Ke\nu)e\nu$ simulation sample (bottom) this cascading contribution is clearly targeted by the selections. Note: The cascading simulation samples here are the same as used in the training of the classifier.

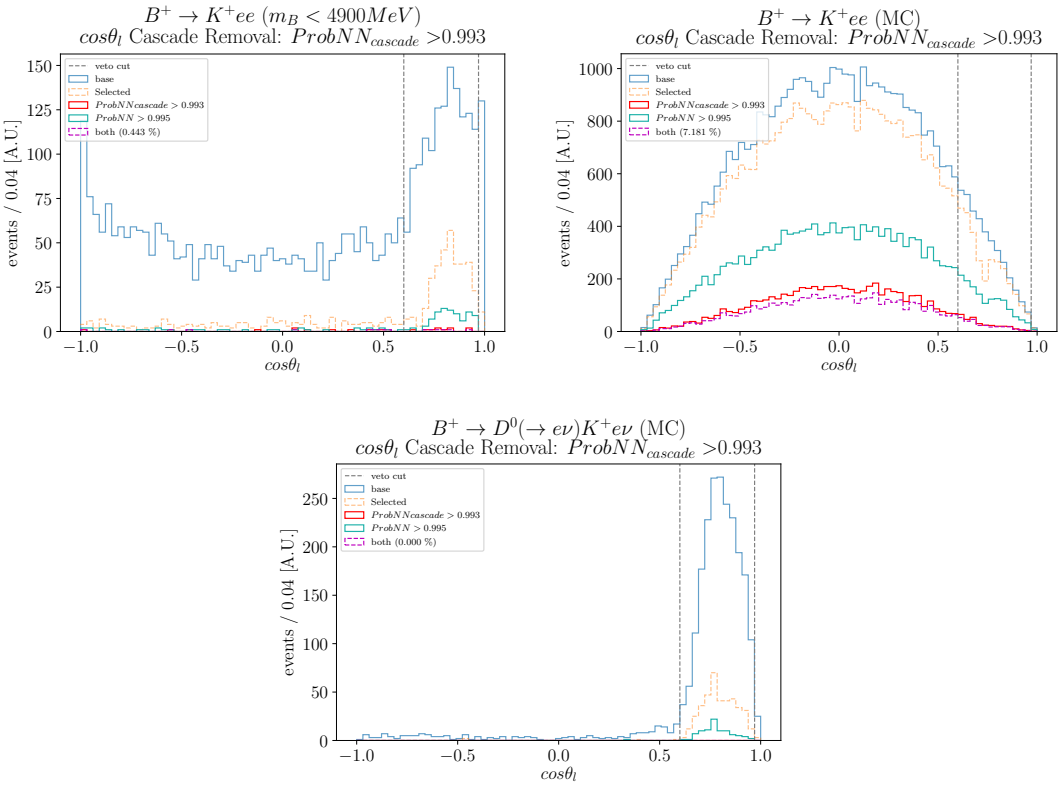

Figure A.35: Studying the iterative rejection efficiency of the focused cuts against the cascading background in the angular distribution for $B^+ \to K^+ee$ data lower mass side band ($m_B < 4900\,\text{MeV}$) (left), $B^+ \to K^+ee$ simulation data (right) and $B^+ \to D^0(\to K^+e\nu)e\nu$ simulation data(bottom). Each plot shows the base starting point in blue, the fully selected in dashed yellow, fully selected with a $\text{ProbNN}_{\text{cascades}}$ cut in red, fully selected with a tighter ProbNN cut in teal, and finally in purple fully selected with a tight ProbNN and $\text{ProbNN}_{\text{cascades}}$ selection applied. The central figure shows the lower mass side band of the $B^+ \to K^+ee$ data, with the error bars displayed, and the $-1.0 < \cos\theta_L < -0.97$ mirrored into the positive reigon, which can be found to be the same magnitude as the upper region.

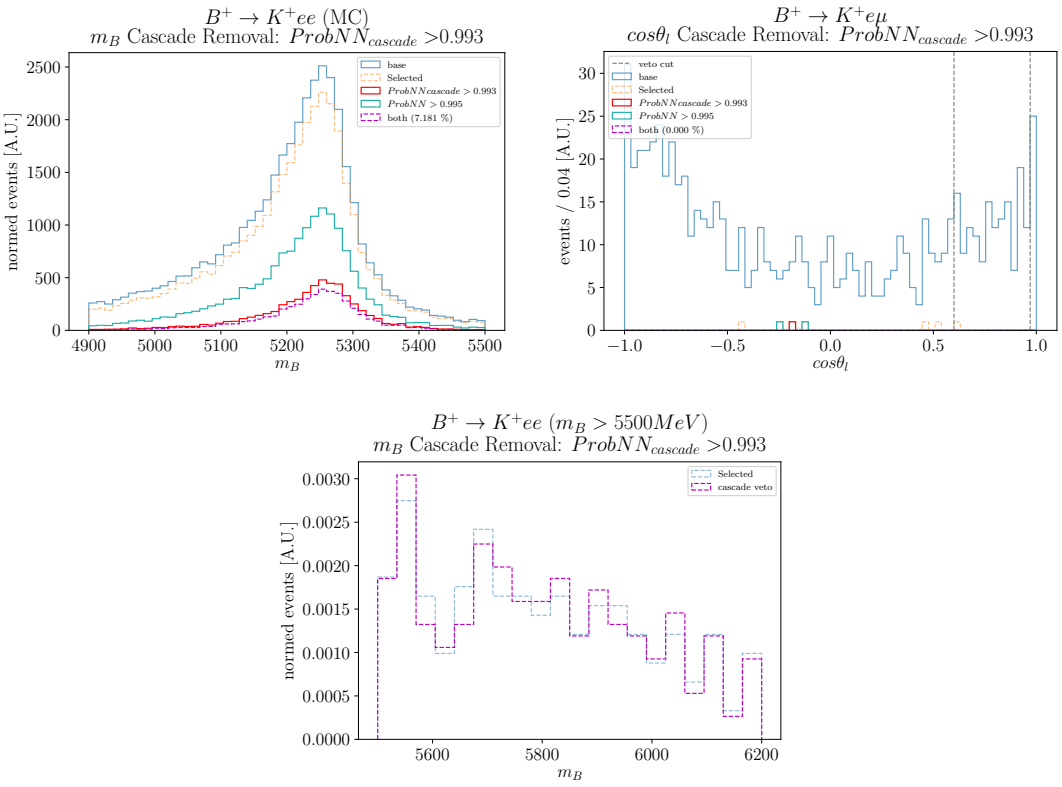

Figure A.36: Studying the iterative rejection efficiency of the focused cuts against the cascading background in the mass distribution for $B^+ \to K^+ee$ simulated samples (left), the angular distribution of $B^+ \to K^+e\mu$ data (right), and $B^+ \to K^+ee$ upper mass side band ($m_B > 5500$ MeV) data (bottom). In the first two plots the colour scheme is the same as in Fig. A.35, in the bottom plot the selected data is shown in blue and the impact of the additional cascade veto shown in magenta. The rightmost plot is normalised to facilitate a comparison of the shape of the background with and without the cut applied to study and sculpting effects, which it can be seen are not present.

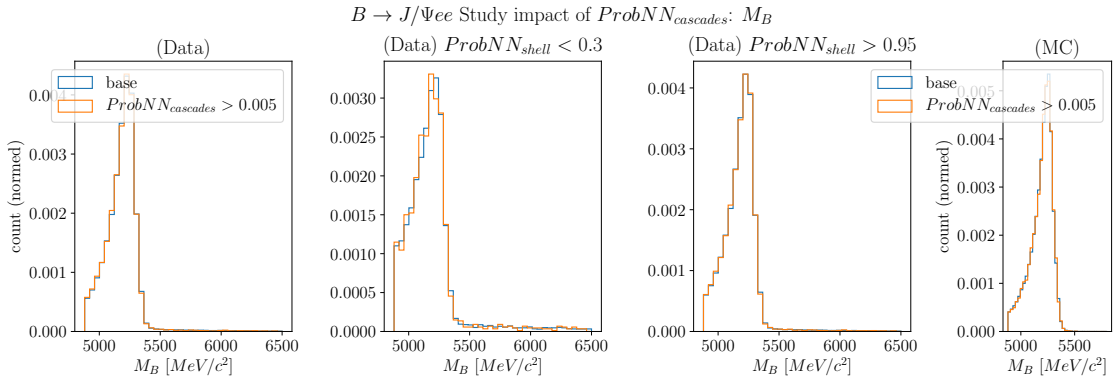

Figure A.37: Studying the sculpting effect of the $\text{ProbNN}_{\text{cascades}} > 0.005$ selection to the reconstructed m_B in $B^+ \to J/\Psi(ee)K^+$ data (left), data in a $\text{ProbNN}_{\text{shell}} < 0.3$ region (center left), data in a $\text{ProbNN}_{\text{shell}} > 0.95$ region, and $B^+ \to J/\Psi(ee)K^+$ simulation data.

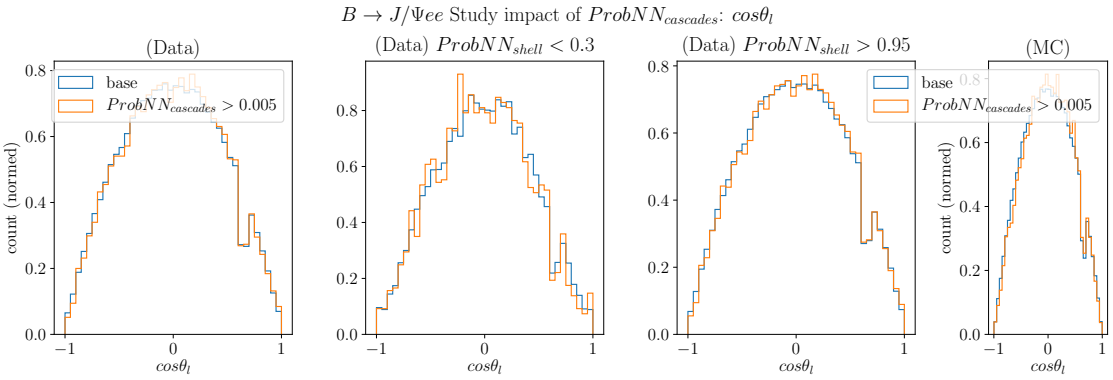

Figure A.38: Studying the sculpting effect of the $\text{ProbNN}_{\text{cascades}} > 0.005$ selection to the reconstructed $cos\theta_l$ distribution for $B^+ \to J/\Psi(ee)K^+$ data (left), data in a $\text{ProbNN}_{\text{shell}} < 0.3$ region (center left), data in a $\text{ProbNN}_{\text{shell}} > 0.95$ region, and $B^+ \to J/\Psi(ee)K^+$ simulation data.

Bibliography

[1] Gianfranco Bertone and Dan Hooper. "History of dark matter". In: *Rev. Mod. Phys.* 90.4 (2018), p. 045002. DOI: 10.1103/RevModPhys.90.045002. arXiv: 1605.04909 [astro-ph.CO].

[2] S. Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (Sept. 2012), pp. 30–61. ISSN: 0370-2693. DOI: 10.1016/j.physletb.2012.08.021. URL: http://dx.doi.org/10.1016/j.physletb.2012.08.021.

[3] **ATLAS Collaboration**. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[4] **LHCb Collaboration**. "Search for lepton-universality violation in $B^+ \to K^+\ell^+\ell^-$ decays". In: (2019). arXiv: 1903.09252 [hep-ex].

[5] Roel Aaij et al. "Angular analysis of charged and neutral $B \to K\mu^+\mu^-$ decays". In: *JHEP* 05 (2014), p. 082. DOI: 10.1007/JHEP05(2014)082. arXiv: 1403.8045 [hep-ex].

[6] A. A. Alves Jr. et al. "The LHCb detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[7] Christoph Bobeth, Gudrun Hiller, and Giorgi Piranishvili. "Angular distributions of $\bar{B} \to \bar{K}\ell^+\ell^-$ decays". In: *JHEP* 12 (2007), p. 040. DOI: 10.1088/1126-6708/2007/12/040. arXiv: 0709.4174 [hep-ph].

[8] Yasmine Sara Amhis et al. "Averages of b-hadron, c-hadron, and τ-lepton properties as of 2018". In: (Sept. 2019). arXiv: 1909.12524 [hep-ex].

[9] Wolfgang Altmannshofer et al. "Symmetries and asymmetries ofB→K∗μμ-decays in the Standard Model and beyond". In: *Journal of High Energy Physics* 2009.01 (Jan. 2009), pp. 019–019. DOI: 10.1088/1126-6708/2009/01/019. URL: https://doi.org/10.1088%2F1126-6708%2F2009%2F01%2F019.

[10] T. Blake, T. Gershon, and G. Hiller. "Rare b Hadron Decays at the LHC". In: *Annual Review of Nuclear and Particle Science* 65.1 (Oct. 2015), pp. 113–143. ISSN: 1545-4134. DOI: 10.1146/annurev-nucl-102014-022231. URL: http://dx.doi.org/10.1146/annurev-nucl-102014-022231.

[11] R Aaij et al. "Measurement of Form-Factor-Independent Observables in the Decay $B^0 \to K^{*0}\mu^+\mu^-$". In: *Phys. Rev. Lett.* 111 (2013), p. 191801. DOI: 10.1103/PhysRevLett.111.191801. arXiv: 1308.1707 [hep-ex].

[12] LHCb collaboration. "Measurement of CP-averaged observables in the $B^0 \to K^{*0}\mu^+\mu^-$ decay". In: *arXiv: High Energy Physics - Experiment* (2020).

[13] T. Hurth, F. Mahmoudi, and S. Neshatpour. "On the new LHCb angular analysis of $B \to K^*\mu^+\mu^-$: Hadronic effects or New Physics?" In: (June 2020). arXiv: 2006.04213 [hep-ph].

[14] Belle Collaboration. *Test of lepton-flavor universality in $B \to K^*\ell^+\ell^-$ decays at Belle*. 2020. arXiv: 1904.02440 [hep-ex].

[15] The LHCb collaboration. "Differential branching fractions and isospin asymmetries of B → K(*)μ+μ decays". In: *Journal of High Energy Physics* 2014.6 (June 2014), p. 133.

[16] Roel Aaij et al. "Measurements of the S-wave fraction in $B^0 \to K^+\pi^-\mu^+\mu^-$ decays and the $B^0 \to K^*(892)^0\mu^+\mu^-$ differential branching fraction". In: *JHEP* 11 (2016). [Erratum: JHEP04,142(2017)], p. 047. DOI: 10.1007/JHEP11(2016)047,10.1007/JHEP04(2017)142. arXiv: 1606.04731 [hep-ex].

[17] Roel Aaij et al. "Angular analysis and differential branching fraction of the decay $B_s^0 \to \phi\mu^+\mu^-$". In: *JHEP* 09 (2015), p. 179. DOI: 10.1007/JHEP09(2015)179. arXiv: 1506.08777 [hep-ex].

[18] Roel Aaij et al. "Differential branching fraction and angular analysis of $\Lambda_b^0 \to \Lambda\mu^+\mu^-$ decays". In: *JHEP* 06 (2015). [Erratum: JHEP09,145(2018)], p. 115. DOI: 10.1007/JHEP09(2018)145,10.1007/JHEP06(2015)115. arXiv: 1503.07138 [hep-ex].

[19] William Detmold and Stefan Meinel. "$\Lambda_b \to \Lambda\ell^+\ell^-$ form factors, differential branching fraction, and angular observables from lattice QCD with relativistic b quarks". In: *Phys. Rev. D* 93 (7 Apr. 2016), p. 074501. DOI: 10.1103/PhysRevD.93.074501. URL: https://link.aps.org/doi/10.1103/PhysRevD.93.074501.

[20] CMS Collaboration. "Angular analysis of the decay B0→K∗0+ from pp collisions at s=8 TeV". In: *Physics Letters B* 753 (2016), pp. 424–448. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2015.12.020. URL: http://www.sciencedirect.com/science/article/pii/S0370269315009685.

[21] Marcel Algueró et al. "Emerging patterns of New Physics with and without Lepton Flavour Universal contributions". In: *Eur. Phys. J. C* 79.8 (2019). [Addendum: Eur.Phys.J.C 80, 511 (2020)], p. 714. DOI: 10.1140/epjc/s10052-019-7216-3. arXiv: 1903.09578 [hep-ph].

[22] The LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08005.

[23] LHCb collaboration. "LHCb magnet: Technical Design Report". In: CERN-LHCC-2000-007 (2000).

[24] LHCb collaboration. "LHCb Tracker Upgrade Technical Design Report". In: CERN-LHCC-2014-001 (2014).

[25] LHCb collaboration. "LHCb VELO Upgrade Technical Design Report". In: CERN-LHCC-2013-021 (2013).

[26] Johannes Albrecht et al. "New approaches for track reconstruction in LHCb's Vertex Locator". In: *EPJ Web of Conferences* 214 (Jan. 2019), p. 01042. DOI: 10.1051/epjconf/201921401042.

[27] LHCb collaboration. "LHCb inner tracker: Technical Design Report". In: CERN-LHCC-2002-029 (2002).

[28] LHCb collaboration. "LHCb outer tracker: Technical Design Report". In: CERN-LHCC-2001-024 (2001).

[29] Roel Aaij et al. "LHCb Detector Performance". In: *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].

[30] LHCb collaboration. "LHCb RICH: Technical Design Report". In: CERN-LHCC-2000-037 (2000).

[31] LHCb collaboration. "LHCb Trigger and Online Technical Design Report". In: CERN-LHCC-2014-016 (2014).

[32] Martın Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[33] Marie-Helene Schune, Francesco Polci, and Martino Borsato. *HOP an additional tool for decays involving electrons*. Tech. rep. LHCb-INT-2015-037. CERN-LHCb-INT-2015-037. Geneva: CERN, Nov. 2015. URL: https://cds.cern.ch/record/2102345.

[34] Muriel Pivk and Francois R. Le Diberder. "sPlot: A statistical tool to unfold data distributions". In: *Nucl. Instrum. Meth.* A555 (2005), pp. 356–369. DOI: 10.1016/j.nima.2005.08.106. arXiv: physics/0402083 [physics.data-an].

[35] Lucio Anderlini et al. *The PIDCalib package*. July 2016. URL: http://cds.cern.ch/record/2202412.

[36] Thibaud Humair, Paula Alvarez Cartelle, and Mitesh Patel. "Test of lepton flavour universality using $B^+ \to K^+ l^+ l^-$ ". In: *LHCb-ANA-2017-042* (June 2017). URL: https://cds.cern.ch/record/2270691.

[37] M. Tanabashi et al. "Review of particle physics". In: *Phys. Rev.* D98 (2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001.

[38] R. Aaij et al. "Test of lepton universality with $B^0 \to K^{*0}\ell^+\ell^-$ decays". In: *JHEP* 08 (2017), p. 055. DOI: 10.1007/JHEP08(2017)055. arXiv: 1705.05802 [hep-ex].

[39] Jerome H. Friedman. "Greedy Function Approximation: A Gradient Boosting Machine". In: *Annals of Statistics* 29 (2000), pp. 1189–1232.

[40] Alex Rogozhnikov. *hepml Python Package*. 2017–. URL: https://github.com/arogozhnikov/hep_ml.

[41] Payam Refaeilzadeh, Lei Tang, and Huan Liu. "Cross-Validation". In: *Encyclopedia of Database Systems*. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pp. 532–538. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_565. URL: https://doi.org/10.1007/978-0-387-39940-9_565.

[42] Diederik P. Kingma and Jimmy Ba. "Adam: A Method for Stochastic Optimization". In: *arXiv e-prints*, arXiv:1412.6980 (Dec. 2014), arXiv:1412.6980. arXiv: 1412.6980 [cs.LG].

[43] Jan Therhaag. "TMVA: Toolkit for multivariate data analysis". In: *AIP Conf. Proc.* 1504.1 (2012). Ed. by Theodore E. Simons and George Maroulis, pp. 1013–1016. DOI: 10.1063/1.4771869.

[44] Tomasz Skwarnicki. "A study of the radiative cascade transitions between the Upsilon-prime and Upsilon resonances". DESY-F31-86-02. PhD thesis. Institute of Nuclear Physics, Krakow, 1986.

[45] James Bourbeau and Zigfried Hampel-Arias. "PyUnfold: A Python package for iterative unfolding". In: *The Journal of Open Source Software* 3.26 (June 2018), p. 741. DOI: 10.21105/joss.00741. URL: https://doi.org/10.21105/joss.00741.

[46] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth.* A462 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[47] G. Bohm and G. Zech. "Statistics of weighted Poisson events and its applications". In: *Nucl. Instrum. Meth. A* 748 (2014), pp. 1–6. DOI: 10.1016/j.nima.2014.02.021. arXiv: 1309.1287 [physics.data-an].

[48] Gary J. Feldman and Robert D. Cousins. "Unified approach to the classical statistical analysis of small signals". In: *Phys. Rev. D* 57 (7 Apr. 1998), pp. 3873–3889. DOI: 10.1103/PhysRevD.57.3873. URL: https://link.aps.org/doi/10.1103/PhysRevD.57.3873.

[49] David M. Straub. "flavio: a Python package for flavour and precision phenomenology in the Standard Model and beyond". In: (2018). arXiv: 1810.08132 [hep-ph].

[50] G. D'Agostini. "Improved iterative Bayesian unfolding". In: *arXiv e-prints*, arXiv:1010.0632 (Oct. 2010), arXiv:1010.0632. arXiv: 1010.0632 [physics.data-an].

[51] LHCb collaboration. "LHCb PID Upgrade Technical Design Report". In: CERN-LHCC-2013-022 (2013).

[52] Roel Aaij et al. "Allen: A high level trigger on GPUs for LHCb". In: *Comput. Softw. Big Sci.* 4.1 (2020), p. 7. DOI: 10.1007/s41781-020-00039-7. arXiv: 1912.09161 [physics.ins-det].

[53] John Nickolls et al. "Scalable Parallel Programming with CUDA". In: *Queue* 6.2 (Mar. 2008), pp. 40–53. ISSN: 1542-7730. DOI: 10.1145/1365490.1365500. URL: https://doi.org/10.1145/1365490.1365500.

[54] Adam Paszke et al. "Automatic differentiation in PyTorch". In: (2017).

[55] Roel Aaij et al. "LHCb Detector Performance". In: *Int. J. Mod. Phys. A* 30.07 (2015), p. 1530022. DOI: 10.1142/S0217751X15300227. arXiv: 1412.6352 [hep-ex].

[56] Roel Aaij et al. "A new algorithm for identifying the flavour of B_s^0 mesons at LHCb". In: *JINST* 11.05 (2016), P05010. DOI: 10.1088/1748-0221/11/05/P05010. arXiv: 1602.07252 [hep-ex].

[57] R Aaij et al. "The LHCb Trigger and its Performance in 2011". In: *JINST* 8 (2013), P04022. DOI: 10.1088/1748-0221/8/04/P04022. arXiv: 1211.3055 [hep-ex].

[58] V.V. Gligorov and Mike Williams. "Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree". In: *JINST* 8 (2013), P02013. DOI: 10.1088/1748-0221/8/02/P02013. arXiv: 1210.6861 [physics.ins-det].

[59] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[60] Phillip Isola et al. "Image-to-Image Translation with Conditional Adversarial Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 5967–5976.

[61] Christian Ledig et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 105–114.

[62] S. Agostinelli et al. "Geant4: A simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[63] John Allison et al. "Geant4 developments and applications". In: *IEEE Trans.Nucl.Sci.* 53 (2006), p. 270. DOI: 10.1109/TNS.2006.869826.

[64] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[65] Jose A Hernando. "The Kalman filter technique applied to track fitting in GLAST". In: (). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.48.1880.

[66] David M. Kreps. "Nash Equilibrium". In: *Game Theory*. Ed. by John Eatwell, Murray Milgate, and Peter Newman. London: Palgrave Macmillan UK, 1989, pp. 167–177. ISBN: 978-1-349-20181-5. DOI: 10.1007/978-1-349-20181-5_19. URL: https://doi.org/10.1007/978-1-349-20181-5_19.

[67] C. Ahdida et al. "Fast simulation of muons produced at the SHiP experiment using Generative Adversarial Networks". In: *JINST* 14 (2019), P11028. DOI: 10.1088/1748-0221/14/11/P11028. arXiv: 1909.04451 [physics.ins-det].

[68] Phillip Isola et al. "Image-to-Image Translation with Conditional Adversarial Networks". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 5967–5976.

[69] Christian Ledig et al. "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 105–114.

[70] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

[71] Nvidia. *NVIDIA TESLA P100 specifications*. https://www.nvidia.com/en-gb/data-center/tesla-p100/. 2020 (accessed 18 Aug, 2020).

[72] Tim Salimans et al. *Improved Techniques for Training GANs*. 2016. arXiv: 1606.03498 [cs.LG].

[73] Inc. Google. *In-Datacenter Performance Analysis of a Tensor Processing Unit*. 2017. arXiv: 1704.04760 [cs.AR].

[74] Kamil Rocki et al. *Fast Stencil-Code Computation on a Wafer-Scale Processor*. 2020. arXiv: 2010.03660 [cs.DC].

[75] Graphcore. *Graphcore.ai*. https://www.graphcore.ai/. 2020 (accessed 24 July, 2020).

[76] Lakshan Ram Madhan Mohan et al. *Studying the potential of Graphcore IPUs for applications in Particle Physics*. 2020. arXiv: 2008.09210 [physics.comp-ph].

[77] Intel. *Intel Xeon Platinum 8168 specifications*. https://ark.intel.com/content/www/us/en/ark/products/120504/intel-xeon-platinum-8168-processor-33m-cache-2-70-ghz.html. 2020 (accessed 18 Aug, 2020).

[78] Graphcore. *private communication*.

[79] Zhe Jia et al. "Dissecting the Graphcore IPU Architecture via Microbenchmarking". In: (2019). arXiv: 1912.03413 [cs.DC].

[80] Dominic Masters and Graphcore. "Delving Deep into Modern Computer Vision Models". In: (2020). https://www.graphcore.ai/posts/introducing-second-generation-ipu-systems-for-ai-at-scale.

[81] N. V. Chawla et al. "SMOTE: Synthetic Minority Over-sampling Technique". In: *arXiv e-prints*, arXiv:1106.1813 (June 2011), arXiv:1106.1813. arXiv: 1106.1813 [cs.AI].